CAMPUS
DE EXCELENCIA
INTERNACIONAL

Graduado en Ingeniería Informática

Universidad Politécnica de Madrid

Facultad de Informática

TRABAJO FIN DE GRADO

Desarrollo de una plataforma de servicios gestionados para soluciones de contingencia en el Cloud

Autor: Manuel Hernández Sánchez

Director: Maria Luisa Córdoba

MADRID, JUNIO DE 2014

RESUMEN

La proliferación en todo el mundo de las soluciones basadas en la nube hace que las empresas estén valorando mover su infraestructura o parte de ella a la nube, para así reducir los altos costes de inversión necesarios para mantener una infraestructura privada. Uno de los servicios que puede ser centralizado en la nube, mediante recursos compartidos entre varios clientes, son las soluciones de contingencia, como los servicios de protección de datos o los centros de recuperación ante desastres.

Mediante este proyecto se pretende llevar a cabo el despliegue de una plataforma de servicios gestionados para ofrecer soluciones centralizadas, a clientes que lo requieran, de copias de seguridad y disaster recovery.

Para la realización del proyecto se realizó un estudio de las tecnologías actuales para llevar a cabo la continuidad de negocio, los distintos tipos de backups, así como los tipos de replicación existente, local y remota. Posteriormente, se llevó a cabo un estudio de mercado para barajar las distintas posibilidades existentes para el despliegue de la infraestructura, siempre teniendo en cuenta el cliente objetivo. Finalmente, se realizó la fase de desarrollo, donde se detallan los componentes principales que componen la solución final, la localización de la infraestructura, un caso de uso, así como las principales ventajas de la solución.

Se ha de destacar que se trata de un proyecto real, que se llevó a cabo en una empresa externa a la facultad, Omega Peripherals, donde una vez finalizado mi prácticum, se propuso dicho proyecto para desarrollarlo como continuación de mi labor en la empresa y formar parte de mi Trabajo Final de Grado (TFG).

ABSTRACT

The worldwide proliferation of cloud-based solutions means that companies are evaluating their infrastructure or move part of it to the cloud, to reduce the high investment costs required to maintain a private infrastructure. One of the services that can be centralized in the cloud, through shared resources between multiple clients, are the solutions of contingency services as data protection or disaster recovery centers.

This project aims to carry out the deployment of a managed services platform centralized solutions, to customers who need it, backup and disaster recovery services.

The project consists of three phases. First, It was performed a study of the current business continuity technologies, the different types of backups, as well as replication types, local and remote. Second, it was performed a market study to shuffle the different possibilities for the deployment of infrastructure, keeping in mind the target customer. Finally, we found the development phase, where it details the main components that make up the final solution, the location of infrastructure, a use case, as well as the main advantages of the solution.

It should be emphasized that this is a real project, which was carried out in an outside company to the university, called Omega Peripherals, where once I completed my practicum, it was proposed this project to develop it as a continuation of my job and develop it as my final dissertation.

ÍNDICE

1 INTRODUCCIÓN

En este documento presento el desarrollo de una plataforma de servicios gestionados para ofrecer soluciones a clientes que lo requieran para sus infraestructuras de contingencia, copias de seguridad y recuperación ante desastres. Dicho trabajo fue desarrollado en la empresa de tecnología de la información, Omega Peripherals, para el proyecto final de Graduado en Ingeniería Informática en la Universidad Politécnica de Madrid.

Hoy en día y cada vez más, las empresas mueven una cantidad muy elevada de datos digitales y es vital salvaguardar toda esa información, para que en caso de algún problema la empresa pueda seguir desarrollando su trabajo con el mínimo impacto posible, por lo que la inversión en soluciones de contingencia está siendo cada vez más demandada. El problema radica en los elevados costes que supone para pequeñas y medianas el montar una infraestructura propia, tanto a nivel de hardware y software, como el personal requerido para administrarla. Además de las reticencias que muestran dichas empresas frente a soluciones existentes de grandes compañías.

El motivo de realizar este proyecto se debe a la necesidad en el mercado actual de una plataforma de contingencia a nivel nacional para dar posibilidad a todas aquellas empresas que quieran y necesiten una infraestructura para salvaguardar sus datos, así como la recuperación de la producción en caso de desastre, manteniendo unos costes mínimos, evitando una gran inversión en infraestructura y en personal. Además de un valor añadido de realizar un proyecto real en una empresa externa a la facultad.

El objetivo fue desarrollar una plataforma óptima que solvente ese problema, realizando un estudio previo de las soluciones actuales de backup y disaster recovery para así obtener una solución lo más balanceada posible en cuanto a rendimiento y costes, para posteriormente montar una infraestructura en un centro de procesamiento de datos, CPD, y por último, obtener un servicio que se comercialice hacia el cliente potencial anteriormente mencionado.

1.1 Objetivos del proyecto

El objetivo del proyecto es proporcionar a las empresas un entorno compartido para desplegar sus soluciones de disaster recovery y copias de seguridad, garantizando el cumplimiento de los **SLA (Service Level Agreement) [1]** o acuerdos de nivel de servicio establecidos.

- Una integración automatizada con el mecanismo de replicación elegido por el cliente.
- Un mecanismo de prueba automatizada de contingencia que asegura de manera permanente que todo el servicio y sus dependencias se encuentran en situación adecuada para ser puestos en marcha sin errores.

La solución comprende hardware más software, pudiendo ser comercializada como un servicio y cumpliendo con las siguientes premisas:

- Multi empresa
    - Los mismos recursos hardware tienen que ser compartidos entre varias empresas, aislados unas con otras de manera lógica.
- Gestión automatizada.
- Gestión de utilización para la facturación.
- Seguridad.
- Escalable.

En la siguiente figura podemos ver un diagrama de la solución propuesta.

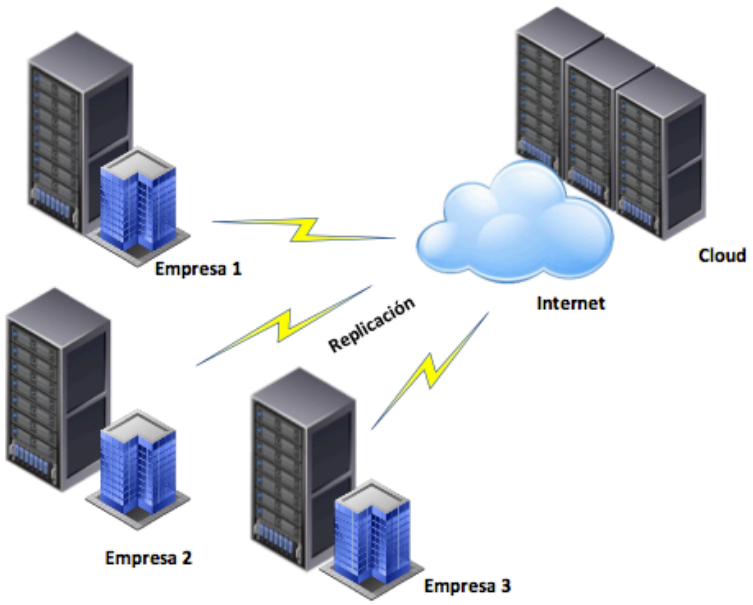

Figura 1. Diagrama alto nivel solución propuesta.

2 CONTINUIDAD DE NEGOCIO

2.1 Introducción a la continuidad de negocio

En el mundo actual, el acceso continuo a la información es una necesidad para el buen funcionamiento de las operaciones de una organización. Los costes y las consecuencias que causa a las empresas la indisponibilidad de la información es mayor que nunca. Hay muchas amenazas contra la disponibilidad de la información, como los desastres naturales, los incidentes no planificados, y los sucesos previstos, que podrían desembocar en inaccesibilidad de la información. . Por lo tanto, es fundamental para las empresas definir estrategias adecuadas que puedan ayudar a superar estas situaciones.

La continuidad del negocio es un proceso importante para definir y poner en práctica estas estrategias. Se define como un proceso que se prepara para, responder y recuperarse ante la caída del sistema que pueda afectar negativamente a las operaciones de la organización. Es decir, la continuidad de negocio es un proceso integrado y extendido a lo largo de toda la empresa, que incluye todas las actividades, tanto internas como externas a ella, que una organización debe realizar para mitigar el impacto del tiempo de inactividad tanto planificado como no planificado. Esto implica la preparación, respuesta y recuperación de una interrupción del sistema que afecta en la producción y en el normal funcionamiento de la organización.

Por tanto, la continuidad de negocio, del inglés business continuity, BC en adelante, consiste en tomar medidas proactivas, como el análisis de impacto en el negocio, la evaluación de riesgos, la implementación de soluciones de tecnología BC (backup y replicación), y contramedidas, como la recuperación de desastres y reinicio de los sistemas, para ser invocadas en el caso de un fallo. El objetivo de una solución BC es asegurar la disponibilidad de la información necesaria para que una organización pueda continuar realizando sus operaciones.

En la siguiente figura se muestra las principales causas de indisponibilidad de la información en las empresas, según datos de la compañía EMC, las cuales se pueden dividir en cortes planeados y no planeados como se ha mencionado anteriormente. Los cortes planeados incluyen la instalación, integración y mantenimiento de hardware, actualización o parches en el software, copias de seguridad, restauración de aplicaciones y datos, así como migraciones al entorno de producción. Los cortes no planeados son los causados por errores humanos, corrupción en la base de datos y fallos en los componentes físicos y/o virtuales de la infraestructura. Otro tipo de incidentes son los desastres naturales o artificiales, tales como fuego, terremotos, entre otros. Como se puede observar en la figura, la mayor parte de los cortes son planeados, seguido de los no planeados y dejando en un porcentaje muy poco probable algún tipo de desastre.

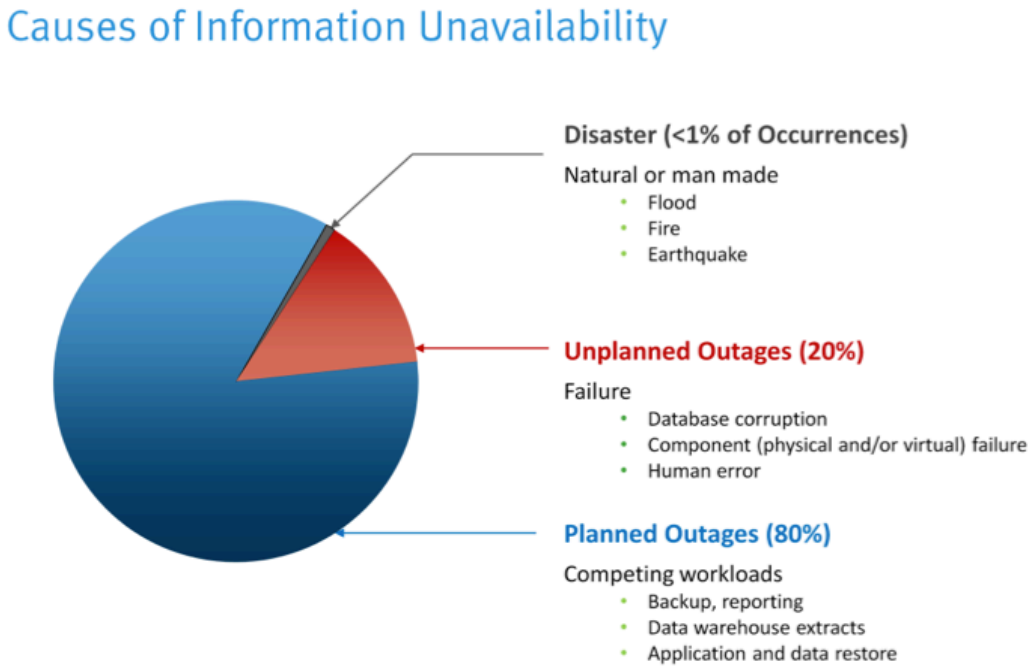

Figura 2. Causas de indisponibilidad de la información. *Copyright 2012 EMC Corporation*

Estas caídas en la infraestructura de las compañías, que provocan la no disponibilidad de la información, causan un tiempo de inactividad en la empresa. Esto da como resultado la pérdida de productividad, pérdida de ingresos y daños, en muchos casos irreparables en su reputación.

El impacto del tiempo de inactividad en una compañía es la suma de todas las pérdidas sufridas como consecuencia de una interrupción determinada. Una medida muy importante que debe tener en cuenta la empresa al afrontar una solución de BC es el coste promedio de inactividad por hora, que proporciona una estimación clave para determinar las soluciones apropiadas de continuidad de negocio. Se calcula de la siguiente forma:

Coste promedio de inactividad por hora

Pérdida de productividad media / h + Pérdida media de ingresos / h

Siendo:

$$\text{Pérdida de productividad / h} = \frac{\text{Total salarios de todos los empleados por semana}}{\text{Media de las horas trabajadas por semana}}$$

$$\text{Pérdida media de ingresos / h} = \frac{\text{Ingresos totales por semana}}{\text{Media de horas de apertura por semana}}$$

La disponibilidad de la información recae en el funcionamiento y disponibilidad tanto de los elementos físicos como virtuales de la infraestructura de la compañía. El fallo de estos componentes podría interrumpir el acceso a la información, provocando las consecuencias anteriormente mencionadas. Para dar solución a un fallo de estos componentes, se puede restaurar mediante un reinicio manual o mediante la reparación o sustitución del sistema afectado. La solución al problema forma parte de las contramedidas anteriormente mencionas, pero en un plan de BC se deben llevar a cabo medidas de análisis de riesgo, donde se toman medidas muy importantes para la BC, como el ratio de fallo de los componentes de la infraestructura y el tiempo medio de reparación, llamadas MTBF y MTTR.

Mean Time Between Failure (MTBF): Es el tiempo medio de un sistema funcionando normalmente entre fallos. Es la medida de la fiabilidad de un sistema o componente y por lo que general se expresa en horas.

MTBF = Total uptime / nº de fallos

Mean Time To Repair (MTTR): Es el tiempo medio requerido para la reparación de un fallo en un componente de la infraestructura. Esta medida incluye el tiempo requerido para: detectar el fallo, avisar al equipo encargado del mantenimiento, diagnosticar el fallo, obtener los componentes necesarios para la reparación, reparar el fallo, probar que se ha reparado y restaurar los datos.

MTTR = Total downtime / nº de fallos

Por tanto la disponibilidad de la información (IA) puede expresarse en términos de el tiempo de disponibilidad de los sistemas (system uptime y system downtime) y puede medirse como la cantidad o porcentaje que un sistema está disponible de la siguiente forma:

IA = System uptime / (system uptime + system downtime)

IA = MTBF / (MTBF + MTTR)

En la siguiente figura, procedente de cursos de formación de la compañía EMC, podemos ver un esquema de la medición de la disponibilidad de la información donde se explica a lo largo del tiempo el tiempo medio requerido para reparar una avería en un componente, es decir, el valor MTTR.

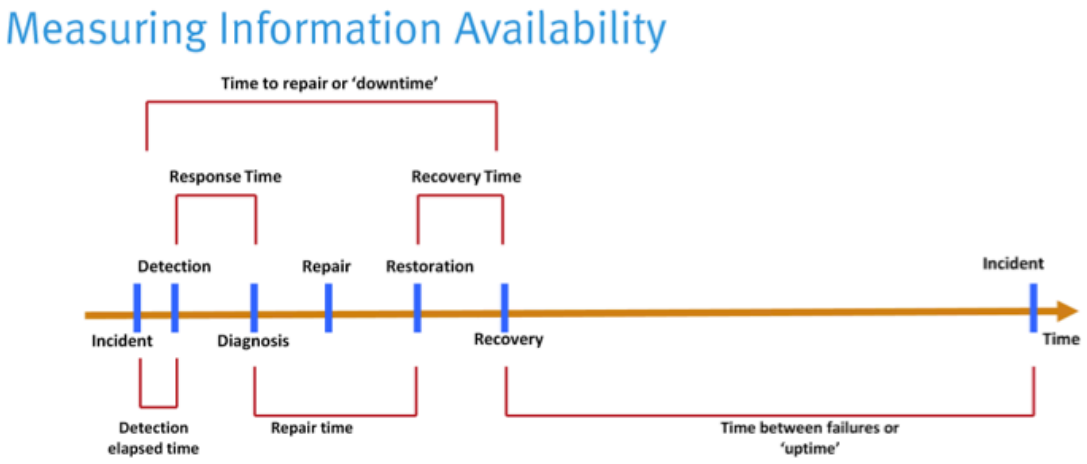

Figura 3. Disponibilidad de la información. *Copyright 2012 EMC Corporation*

Como se ha comentado, la disponibilidad puede expresarse mediante un porcentaje que representa la disponibilidad al año de un sistema. La siguiente tabla muestra una lista con los porcentajes típicos de disponibilidad. Se ha de mencionar el hecho de que un sistema esté disponible, no significa que se pueda acceder a él, pudiendo existir, por ejemplo, algún problema en la red que nos impida hacer uso del mismo. Este porcentaje mide exactamente la disponibilidad al año cuando el sistema es accesible y disponible.

Uptime (%)	Downtime (%)	Downtime por año	Downtime por semana
98	2	7.3 días	3 hrs, 22 min
99	1	3.65 días	1 hr, 41 min
99.8	0.2	17 hrs, 31 min	20 min, 10 sec
99.9	0.1	8 hrs, 45 min	10 min, 5 sec
99.99	0.01	52.5 min	1 min
99.999	0.001	5.25 min	6 sec
99.9999	0.0001	31.5 sec	0.6 sec

Tabla 1. Tiempos y porcentajes de disponibilidad.

Otro punto a tratar cuando hablamos de continuidad de negocio es el concepto de recuperación de desastre, **disaster recovery** y de restaurado del desastre, **disaster restart**.

Disaster recovery es el proceso coordinado de la restauración de los sistemas, los datos y la infraestructura necesaria para sostener las operaciones en curso, después de un desastre que haya interrumpido el entorno de producción de la empresa. Es decir, es el proceso de restauración de la última copia realizada, aplicando los logs necesarios para volver a un punto de consistencia conocido. Una vez todo el proceso de recuperación se ha completado, los datos son validados para asegurar que son correctos.

Disaster restart es el proceso de reinicio de las operaciones críticas, para que la compañía pueda seguir operando, llevada a cabo mediante replicación de los datos y aplicaciones.

Una vez conocemos estos dos términos, podemos definir dos mediciones muy importantes cuando hablamos de BC, que son el punto objetivo de recuperación (RPO) y el tiempo objetivo de recuperación (RTO).

Recovery-Point Objetive (RPO) es el punto en el tiempo en que los sistemas y los datos deben ser recuperados después de un corte en el servicio. Define la cantidad de pérdida de datos que una empresa puede soportar. Basándose en la RPO, las organizaciones pueden planifican la frecuencia con la que se deben hacer copias de seguridad o réplicas. Por lo tanto, a menor RPO, se dispone de copias más recientes, puesto que se realizan mayor cantidad de réplicas o backup de la información.

Recovery-Time Objetive (RTO) es el tiempo que tardan los sistemas y las aplicaciones en volver a la normalidad después de una caída en el servicio. Define la cantidad de tiempo de inactividad en las operaciones críticas que una empresa puede soportar.

En la siguiente figura podemos ver las diferentes tecnologías usadas según el tiempo RPO y RTO que podemos obtener con cada una de ellas.

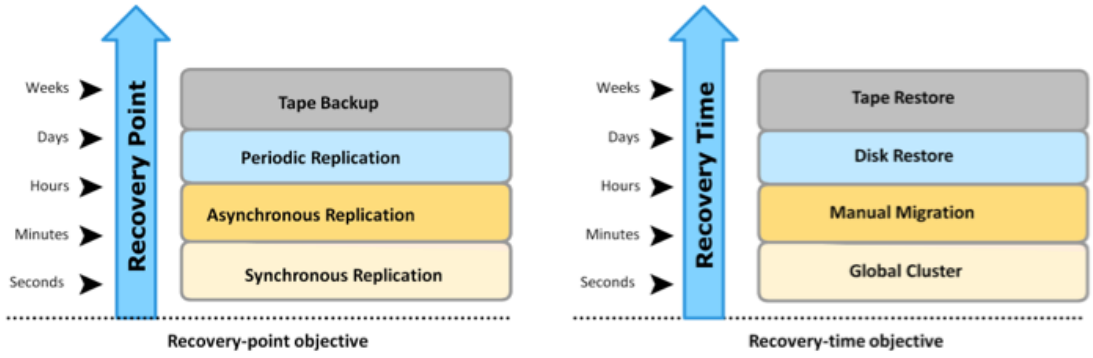

Figura 4. Tecnologías según RPO y RTO

2.1.1 Planificación de BC

La planificación de la continuidad de negocio debe seguir un enfoque estructurado y ordenado, como cualquier otro proceso de planificación. Actualmente, las organizaciones dedican muchos recursos para desarrollar y mantener un plan de BC. El ciclo de vida que debe tener el plan, debe incluir las siguientes cinto etapas, como se muestra en la siguiente figura, según los cursos de formación de EMC.

Figura 5. Ciclo de vida BC.*Copyright 2012 EMC Corporation*

1. **Establecimiento de objetivos:** En este punto se deben determinar los requisitos del plan, para así fijar el alcance y el presupuesto para cumplir dichos requisitos. También se debe seleccionar un equipo de expertos, así como determinar las políticas necesarias.

2. **Análisis:** Recopilación de información de perfiles de datos, procesos del negocio, soporte y frecuencia de uso de la infraestructura y dependencias. Se ha de identificar los procesos críticos, analizar los riesgos, asignar prioridades y crear estrategias de mitigación. Por último se deben realizar análisis de coste-beneficio para dichas estrategias, evaluando las diferentes opciones.

3. **Diseño y desarrollo:** En este paso se debe definir la estructura del equipo y asignar funciones y responsabilidades individuales, así como diseñar estrategias de protección de datos, solución de contingencia y procedimientos de respuesta de emergencia, siendo de vital importancia detallar los procedimientos de recuperación y reinicio.

4. **Implementación:** Implementar procedimientos de gestión de riesgos y mitigación que incluyen backup, replicación y gestión de los recursos, así como la preparación del sitio destinado a disaster recovery, para actuar en caso de que un desastre afecte el centro de datos primario. Además es necesario evitar los puntos únicos de fallo, redundando aquellos sistemas necesarios.

5. **Formación, pruebas, evaluación y mantenimiento:** En este último paso se ha de entrenar y formar al equipo responsable regularmente, o cuando se realiza alguna modificación en el plan, mostrando los procedimientos de respuesta y de recuperación. Además se ha de llevar a cabo los procesos de evaluación de daños y revisar los planes de recuperación para mantenerlos al día, así como realizar pruebas regularmente para evaluar el buen desempeño e identificar posible problemas, para posteriormente realizar informes de rendimiento y actuación y poner solución a dichos problemas.

Una vez se ha realizado el plan y se ha puesto en marcha, se realiza un análisis de impacto, llamado BIA, **Business Impact Analysis**, que identifica qué operaciones y procesos son esenciales para la supervivencia de la compañía. Estudia los impactos financieros, operacionales y de servicio que puede provocar una interrupción en los procesos críticos de las empresa, evaluando distintas áreas para determinar la capacidad de recuperación de la infraestructura auxiliar.

Por tanto, con este análisis se puede producir un informe muy detallado de los incidentes ocurridos y el impacto que puede provocar sobre la compañía, especificado en términos de dinero o en tiempo. Consecuentemente, con un buen análisis, las compañías pueden priorizar e implementar las contramedidas necesarias para mitigar los impactos de una caída en el servicio, que deben estar detalladas en el plan de BC.

Una vez introducidos los conceptos más importantes en la continuidad de negocio, podemos afirmar que un buen estudio y planificación, así como la realización de análisis de impacto, son vitales para las compañías para diseñar e implementar una solución de contingencia de acuerdo a sus necesidades.

2.2 Backup

Una copia de seguridad es una copia adicional de los datos de producción, creado y mantenido con el único propósito de recuperar los datos perdidos o dañados. Con el aumento de las exigencias empresariales y regulatorias para el almacenamiento de datos, la retención y la disponibilidad, las organizaciones se enfrentan a la tarea de backup de una cantidad cada vez mayor de datos. Esta tarea se hace más difícil con el crecimiento de la información, los presupuestos cada vez más reducidos, así como menos tiempo para realizar las copias de seguridad. Por otra parte, las organizaciones necesitan una restauración rápida de la copia de seguridad de datos para cumplir con los acuerdos de nivel de servicio (SLA).

Los backups se pueden realizar para hacer frente a las necesidades de recuperación de desastres, así como para restaurar los datos en un sitio alternativo cuando el sitio primario se encuentra incapacitado debido a un desastre, teniendo en cuenta el RPO y el RTO requeridos, utilizando diferentes estrategias de protección de datos para la recuperación de desastres según las necesidades de cada organización.

Cada día se generan mayor cantidad de datos, por lo que los backups son de vital importancia en el día a día de una empresa para poder restaurar los datos en caso de pérdida de datos o error lógico en la infraestructura. Por ejemplo, es común que un usuario borre accidentalmente un correo electrónico importante o que un archivo se corrompa, por lo que se recurre al backup para su restauración.

También se utilizan para hacer frente a los requisitos de archivado, por las pequeñas y medianas empresas en su mayoría, para la conservación a largo plazo de aquellos datos, como registros de transacciones, mensajes de correo electrónico y demás datos que requieren un archivado con una antigüedad de determinada para el cumplimiento normativo.

2.2.1 Tipos de backup

Existen distintos tipos de backup dependiendo de las necesidades de cada compañía y sus requisitos RTO / RPO. Estos tipos determinan cuándo y de qué se realiza la copia y se pueden clasificar en completo, incremental y acumulativo (o diferencial). La mayoría de las organizaciones utilizan una combinación de estos tres tipos de copia de seguridad para satisfacer sus necesidades. En la siguiente figura, obtenida de los cursos oficiales de la compañía EMC, se puede observar el funcionamiento de cada uno.

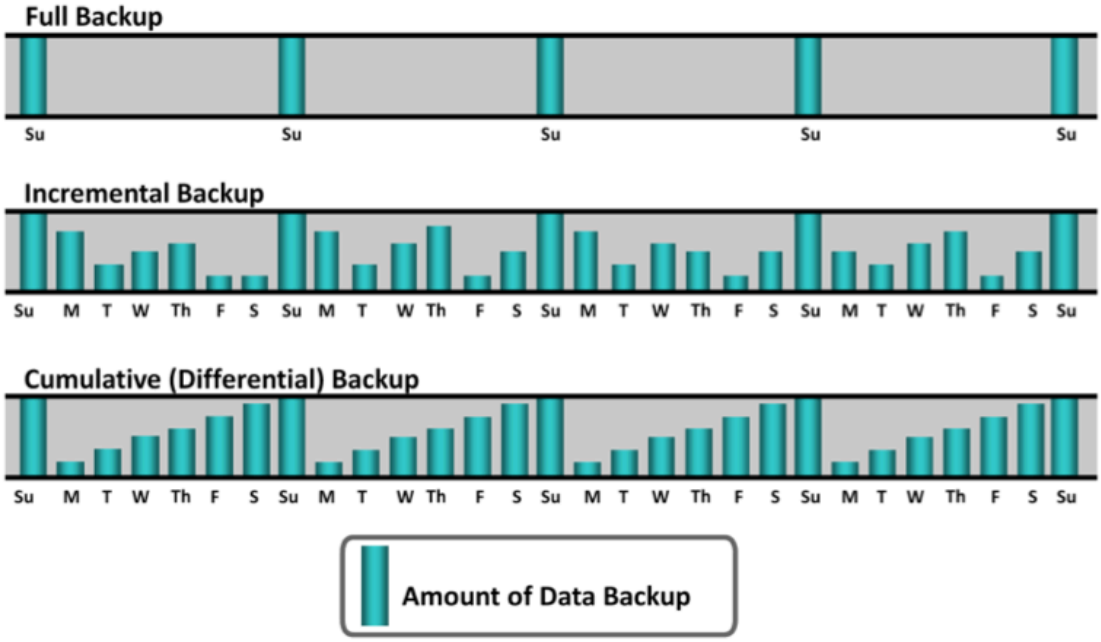

Figura 6. Diagrama de tipos de backup

El backup **completo** es una copia integral de los datos, copiando la totalidad de los contenidos de los sistemas a mantener sobre sistemas de almacenamiento. Proporciona una recuperación más rápida , pero requiere más espacio de almacenamiento y también toma más tiempo en completarse.

En los **incrementales** se copian los datos que han cambiado desde la última copia de seguridad completa o incremental. Este tipo de backups es mucho más rápido que una copia de seguridad completa, debido a que el volumen de datos copiados se limita sólo a aquellos datos que han sido modificados recientemente, pero necesita más tiempo para la restauración.

Por último, en los backups **acumulativos** se copian los datos que han cambiado desde la última copia de seguridad completa. Este método necesita más tiempo que una copia de seguridad incremental , pero hace que la restauración sea mucho más rápida.

Otra forma de realizar un backup completo es mediante una copia de seguridad **sintética** (o artificial) . Este método se utiliza cuando los recursos de producción no se pueden reservar en exclusiva para un proceso de copia de seguridad durante períodos prolongados para realizar una copia completa. Se denomina sintética porque no se crea directamente de los datos de producción, permitiendo una copia completa que se creará sin interrumpir la operación de E / S en el entorno de producción, liberando ancho de banda en la red.

El proceso de restauración desde una copia de seguridad incremental requiere el último backup completo y todos los incrementales disponibles hasta el punto de restauración, lo que perjudica gravemente el tiempo de recuperación al tener que recuperar todas copias realizadas. En la siguiente figura podemos ver un ejemplo de restauración incremental.

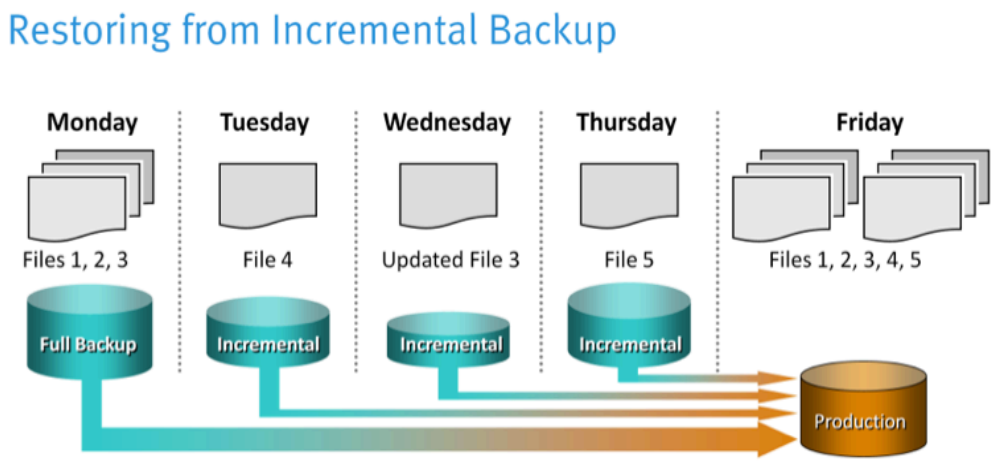

Figura 7. Restauración desde backup incremental *Copyright 2012 EMC Corporation*

El proceso de restauración desde una copia de seguridad diferencial, requiere el último backup completo y el último acumulativo disponible hasta el punto de restauración, de esta manera, se acelera el proceso de recuperación. En la siguiente figura podemos ver un ejemplo de restauración diferencial.

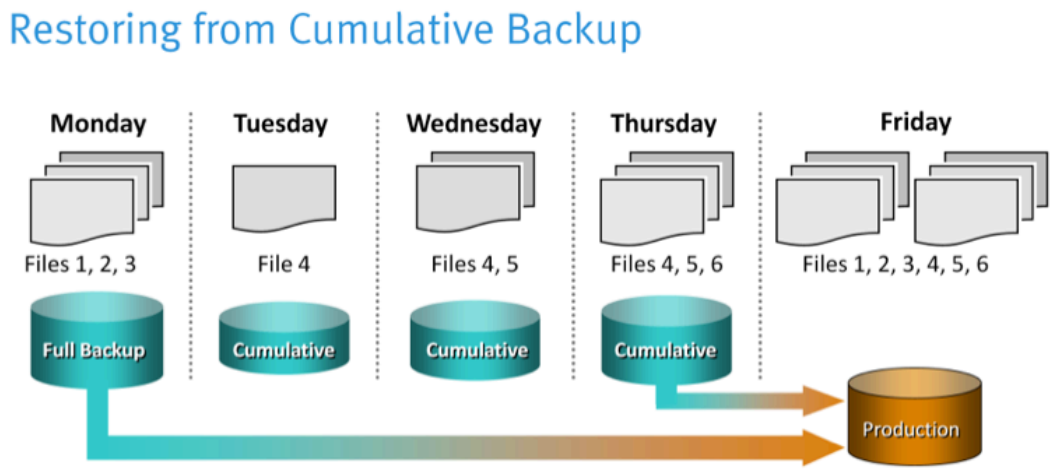

Figura 8. Restauración desde backup acumulativo. *Copyright 2012 EMC Corporation*

2.2.2 Arquitectura de backup

Un sistema de copia de seguridad utiliza comúnmente la arquitectura de cliente - servidor con un servidor de backup y varios clientes. El servidor administra las operaciones de backup y mantiene el catálogo de las copias realizadas, además contiene información sobre la configuración del backup sobre cuándo ejecutar el proceso y qué datos del cliente han sido copiados. El cliente ha de recopilar los datos que han de ser salvaguardados y los mandará al nodo de almacenamiento, además de informar al servidor de backup. En la siguiente figura podemos ver un diagrama de la arquitectura.

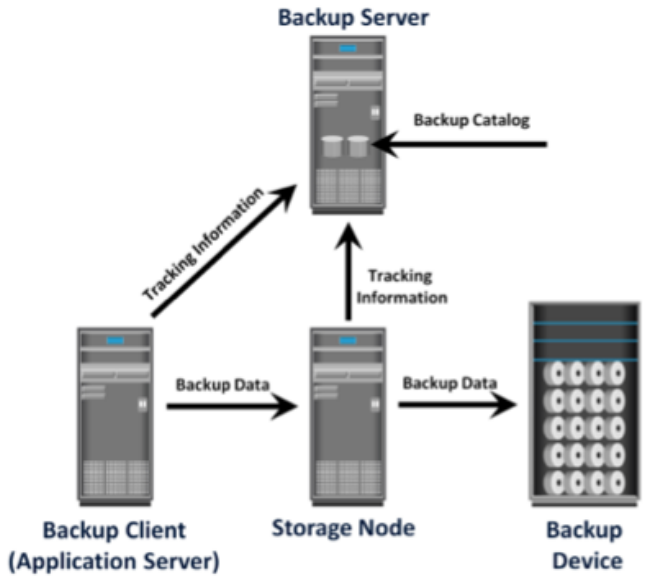

Figura 10. Arquitectura típica backup. *Copyright 2012 EMC Corporation*

El nodo de almacenamiento es responsable de escribir los datos al sistema de backup, además de enviar la información necesaria al servidor. En muchos casos, el nodo de almacenamiento se integra físicamente con el servidor de backup. El dispositivo de backup (disco o librería de cintas) se conecta directamente a través de diferentes tecnologías, que veremos más adelante, al nodo de almacenamiento.

Existen dos tipos de backup según el estado en el que se encuentres los datos en el momento de la copia, el backup en caliente y en frío. En un backup en caliente los datos están siendo usados, con los usuarios accediendo a ellos en el momento de la copia, lo que provoca un impacto en el sistema, debido a los recursos necesarios para realizar la copia, además se incrementan las posibilidades de corrupción de datos. En un backup en frío los datos no están siendo utilizados en el momento de la copia, lo que mejora la consistencia de los datos y reduce las posibilidades de error, pero mientras la copia se lleva a cabo, los datos son inaccesibles para los usuarios.

2.2.3 Topologías de backup

Una vez conocemos la arquitectura típica de un sistema de copia de seguridad y sus componentes principales, podemos definir las diferentes tecnologías existentes para su interconexión que son: backup de conexión directa, vía LAN, vía SAN, mixto y vía NAS. Todas las imágenes que se muestran a continuación pertenecen a los cursos oficiales de la compañía EMC.

En un **backup de conexión directa**, el nodo de almacenamiento se configura en el cliente, y el destino del backup está conectado directamente a él. Sólo los metadatos se envían al servidor de backup a través de la LAN, por lo que se libera la red local de tráfico de backup. A medida que el medio ambiente crece, habrá una necesidad de una gestión centralizada y la distribución de los dispositivos de copia de seguridad para optimizar los costos. Este tipo de soluciones de conexión directa no son la solución ideal, puesto que, al contrario que las basadas en red, ya sea local o de almacenamiento, optimizan la utilización de los dispositivos de backup, ya que estos son compartidos por varios clientes en la infraestructura, lo que reduce los costes y la gestión.

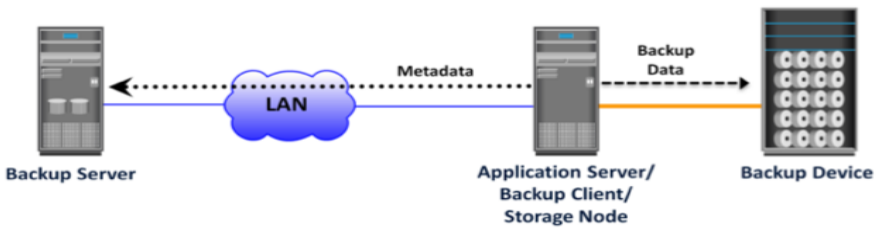

Figura 11. Backup de conexión directa. *Copyright 2012 EMC Corporation*

En el **backup basado en LAN**, los clientes, el servidor, el nodo de almacenamiento y los dispositivos donde se alojarán las copias están conectados a la LAN. Los datos se transfieren desde el cliente al destino a través de la red local, lo que puede provocar una bajada de rendimiento en la red, que se puede minimizar mediante la configuración de redes separadas para backup y la instalación de nodos dedicados para algunos servidores.

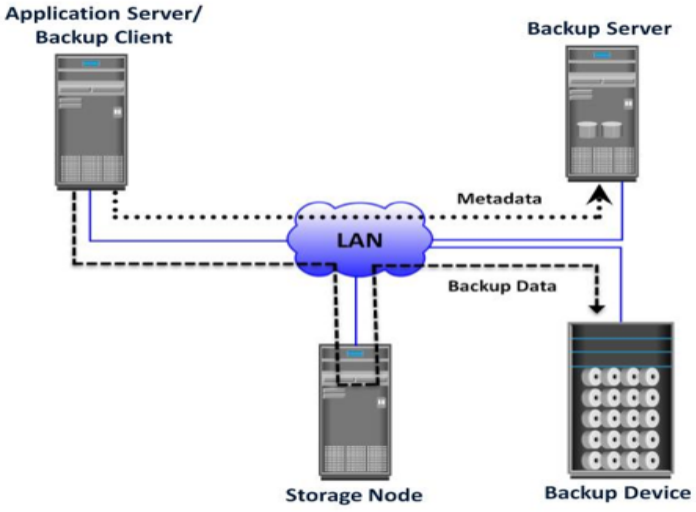

Figura 12. Backup basado en LAN. *Copyright 2012 EMC Corporation*

El **backup basado en SAN** es la solución más adecuada cuando un dispositivo debe ser compartido entre los clientes de la infraestructura, y se conecta a los clientes mediante SAN. Por lo tanto, el tráfico de los datos se limita a la SAN, y sólo los metadatos se transportan a través de LAN, por lo que el rendimiento de la red local no se ve afectado con esta solución.

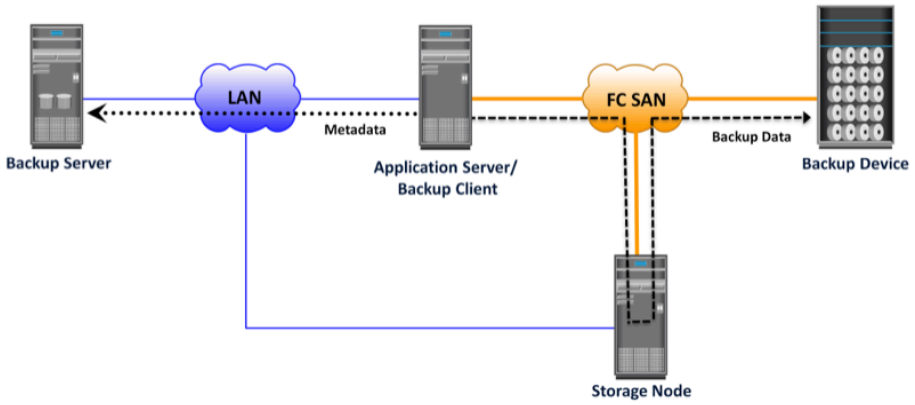

Figura 13. Backup basado en SAN. *Copyright 2012 EMC Corporation*

La topología **mixta** utiliza tanto conexiones de tipo LAN como SAN. Este tipo de solución se suele implementar por razones de coste, ubicación del servidor y como mejora del rendimiento, ya que se usa tanto la red local como la red de almacenamiento para la transferencia de los datos.

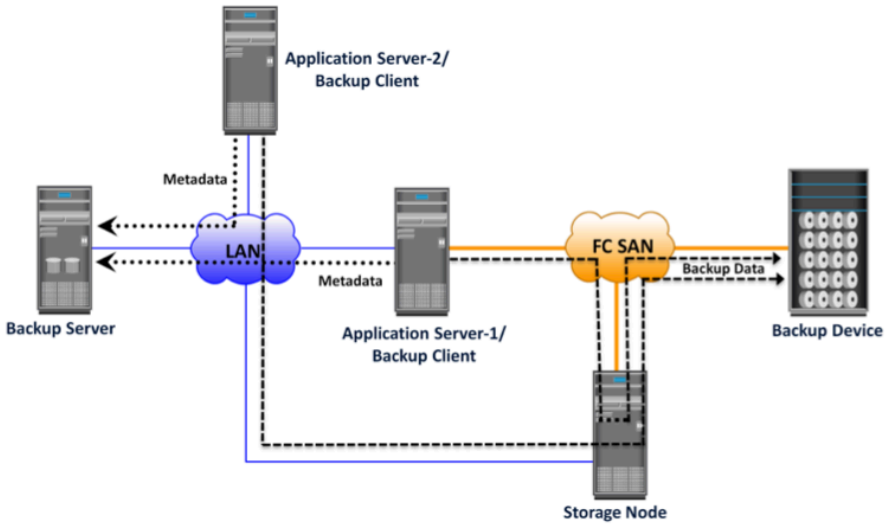

Figura 14. Backup de topología mixta *Copyright 2012 EMC Corporation*

Por otra parte, el uso de los dispositivos de tipo NAS ha hecho que se replantee la estrategia de backup. Estos dispositivos utilizan un sistema operativo en el propio sistema, además soportan múltiples protocolos de intercambio de ficheros En este tipo de soluciones, las copias se pueden implementar de diferentes maneras: mediante un servidor, sin servidor, o con el uso del protocolo Network Data Management Protocol (NDMP).

El protocolo NDMP es un estándar basado en el protocolo TPC/IP específicamente diseñado para la realización de backup en un entorno NAS. Este se comunica con la mayoría de elementos de la infraestructura de backup para realizar la transferencia de archivos para la copia de seguridad. Además, se puede usar dicho protocolo para realizar el backup sin importar la plataforma de origen. Debido a la flexibilidad que aporta NDMP, no es necesario transportar los datos a través del servidor de aplicaciones, lo que reduce la carga del mismo y las tareas de backup se realizan en menor tiempo. Esto se consigue porque permite la conexión directa y el intercambio de datos mediante la cabeza NAS y el dispositivo de backup, mientras que los metadatos se envían al servidor de backup.

En la siguiente figura podemos ver una infraestructura basada en NAS con el uso del protocolo NDMP de dos vías. Como se puede observar, en esta solución el dispositivo de backup se conecta directamente a la cabeza NAS, por lo que no soporta la gestión centralizada de dispositivos de backup, lo que puede suponer un problema a la hora de elegir esta solución, puesto que cada cabeza NAS gestiona el dispositivo al que está conectada.

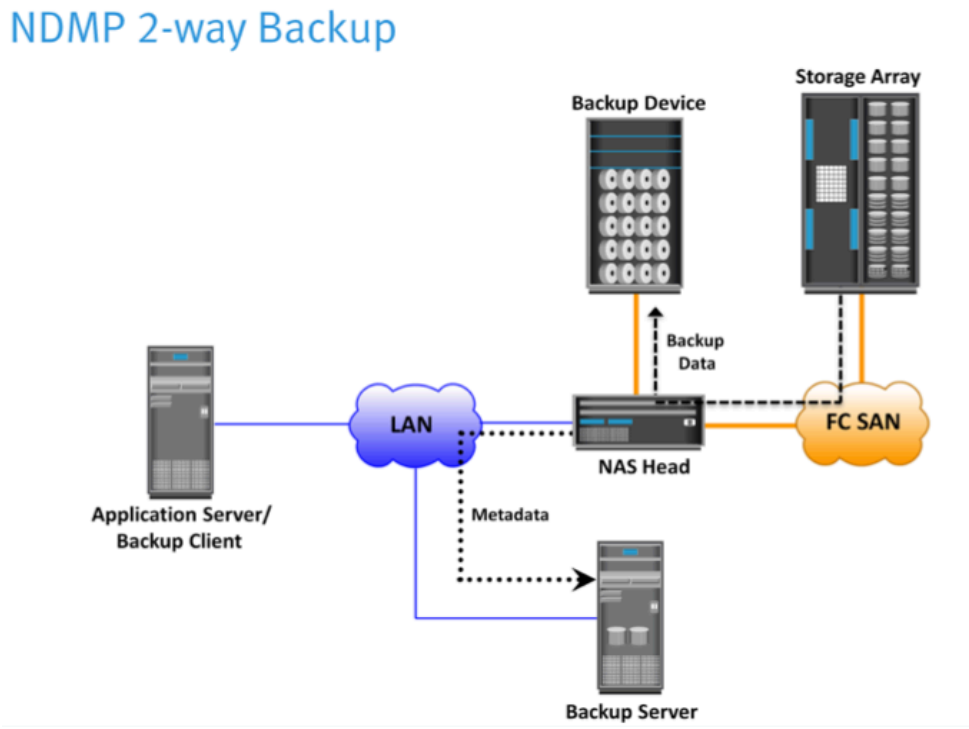

Figura 15. Backup mediante NDMP de 2 vías

Para solucionar el problema de la solución anterior, se implementa otro tipo de topología, también basada en NAS pero mediante NDMP de tres vías, lo que introduce redes separadas para el backup que conectan todas las cabezas NAS, además de al menos unas de ellas conectada al servidor de backup para el envío de metadatos, como se puede ver en la siguiente figura. Este tipo de topología es útil cuando el servidor de backup tiene que estar compartido entre las cabezas NAS, así estas pueden controlar el dispositivo de backup al que estén conectadas.

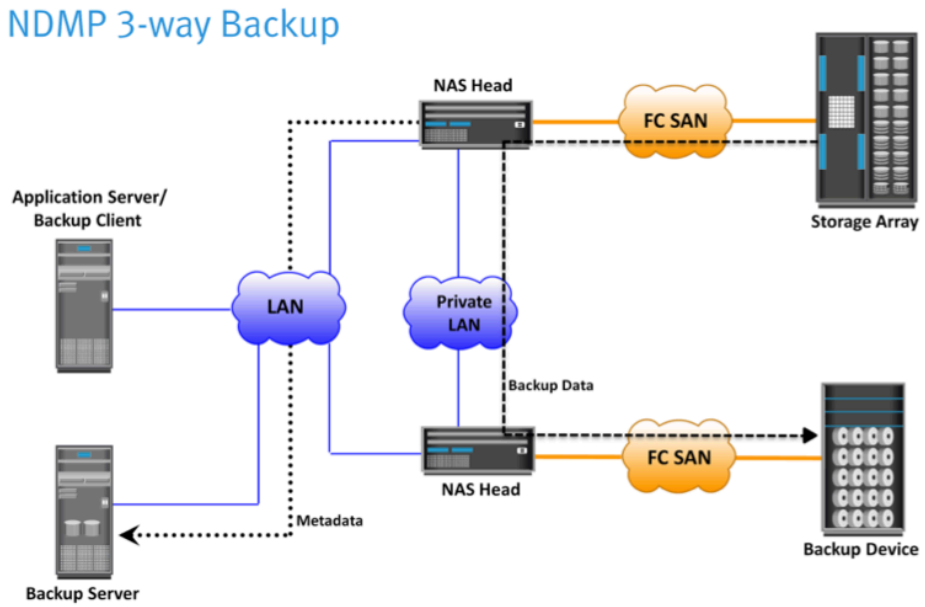

Figura 16. Backup mediante NDMP de 2 vías .*Copyright 2012 EMC Corporation*

2.2.4 Tipos de dispositivos de backup

Se ha mencionado en multitud de ocasiones a lo largo de este documento los dispositivos de backup, que no son más que el lugar donde va a parar la copia de seguridad para almacenamiento. Actualmente se usan tres tipos como almacenamiento, que son: mediante cinta, mediante disco y mediante cintas virtuales.

El **backup a cinta** es el método más extendido, debido en parte a su bajo coste. Para llevar los datos a la cinta, es necesario montarla y leerla, y este proceso se realiza mediante drives. Es un procesos totalmente mecánico donde se introducen las cintas en la librería de cintas, se montan a un drive para escribir o leer los datos y así secuencialmente se van grabando sobre las cintas necesarias hasta completar la tareas. Con este método, las empresas pueden extraer las cintas para guardarlas en un lugar seguro o durante largos periodos de tiempo según normativa. Este método tiene ciertas limitaciones, ya que al ser secuencial y mecanizado el proceso de copia y restaurado son demasiado lentos, por ello, están cayendo en desuso.

El **backup a disco** soluciona estas limitaciones, por lo que gracias a que el coste de los discos ha ido disminuyendo, cada vez hay más compañías que optan por este tipo de solución, debido a las ventajas que suponen respecto a las cintas en cuanto a rendimiento, facilidad de implementación, reducción de tiempos en copia y restauración, así como una mejora en la calidad del servicio.

Tabla 2

En muchas empresas se opta por una solución mixta, usando el backup a disco temporalmente para después pasar la copia o parte de ella a cinta y poder guardarlas durante largos periodos por razones legales. Esto se hace para mejorar el rendimiento del sistema sin dejar de lado los beneficios que suponen el uso de cintas.

Por último tenemos el **backup mediante cintas virtuales**, que no es más que discos que mediante software emulan el funcionamiento de una librería de cintas, simulando desde el mecanismo de montado y lectura hasta los códigos de barras que identifican cada cinta. Este sistema ofrece grandes beneficios, puesto que al emular el funcionamiento de las cintas, se pueden exportar datos preservando el código de cada cinta, lo que facilita la exportación a cinta física y la restauración de las mismas.

En la siguiente tabla podemos ver un resumen de los tipos de dispositivos disponibles, comparados según sus características.

	Cinta	Disco	Cinta Virtual
Capacidad de replicación	NO	SI	SI
Fiabilidad	No dispone de métodos internos para garantizarla	RAID, spare	RAID, spare
Rendimiento	Bajo	Alto	Alto
Uso	Sólo Backup	Multiple	Sólo Backup

Tabla 2. Tipos de dispositivos según características.

2.2.5 Deduplicación

Las soluciones de backup tradicionales no proporcionan ningún mecanismo o herramienta implícita para evitar la duplicación de datos, por lo que se realizan copias de gran cantidad de datos duplicados. Esto provoca que se aumente significativamente el uso de recursos para realizar el backup, el espacio necesario de almacenamiento, así como el ancho de banda, aumentado el tamaño de la ventana de backup. Con el gran crecimiento de la información y los requisitos de disponibilidad de aplicaciones cada vez mayores, las ventanas de backup se están reduciendo considerablemente, lo que provoca un conflicto con las soluciones tradicionales al penalizar el tiempo de backup.

La deduplicación es el proceso de identificar y eliminar los datos redundantes. Cuando se detectan datos duplicados durante la copia de seguridad, estos se descartan y sólo se crea el puntero para referenciar la copia que ya fue respaldada. La deduplicación es imprescindible para reducir los requisitos de almacenamiento, reducir la ventana de backup y eliminar la sobrecarga de la red, lo que permite almacenar más copias en el disco y conservar los datos durante más tiempo.

Existen dos métodos de deduplicación, **a nivel de archivo** y **a nivel de bloque**. Ambos métodos ofrecen buenos resultados, aunque estos pueden variar dependiendo del entorno en el que se usan. Las diferencia es la cantidad de reducción de datos que cada método consigue, así como el tiempo que cada uno toma para analizar los datos y procesarlos para encontrar duplicados.

También podemos catalogar la deduplicación según dónde son analizados los datos y filtrados para eliminar la redundancia, pudiendo ser deduplicación **en origen** o **en destino**.

La deduplicación **a nivel de archivo** detecta y elimina las copias redundantes de los archivos idénticos. Permite almacenar una sola copia del archivo, para en las copias posteriores sustituir por un puntero que apunta al archivo original. Es un método sencillo y rápido, pero no soluciona el problema de contenido duplicado dentro de los archivos.

La deduplicación **a nivel de bloque** rompe el archivo en partes más pequeñas y luego mediante un algoritmo detecta los datos redundantes a nivel de bloque de datos. Como resultado, este método elimina mayor cantidad de datos duplicados al actuar a un nivel más bajo. Hay dos formas de deduplicación a nivel de bloque: **bloques de longitud fija** y **segmentos de longitud variable**. El primero divide los archivos en bloques de longitud fija y utiliza un algoritmo hash para encontrar los datos duplicados. Se trata de un método simple, pero no muy eficiente, ya que al fijarse en bloques de un determinado tamaño, si un archivo cambia o se modifica en su interior añadiendo alguna línea, por ejemplo el nombre del autor, el bloque cambia completamente aunque el archivo sigue siendo el mismo, por lo que se detectaría la duplicación. En segmentos de longitud variable, si hay un cambio en el segmento, el límite para ese segmento se ajusta, dejando los segmentos restantes sin cambios. Este método mejora considerablemente la capacidad de encontrar segmentos de datos duplicados en comparación con el de bloque fijo, lo que hace que la deduplicación sea más eficiente.

Cuando la deduplicación se realiza **en origen**, se eliminan los datos redundantes antes de que se transmitan al destino, lo que permite reducir drásticamente la cantidad de datos enviados por la red durante los procesos de backup, reduciendo la ventana de backup, requiriendo menos ancho de banda en la red, así como menor espacio de almacenamiento para almacenar la copia. Por otro lado, este tipo de deduplicación aumenta la carga en el cliente de backup perjudicando el rendimiento del sistema origen. Además, no todas las aplicaciones de backup soportan este tipo de deduplicación en origen, lo que puede conllevar sobrecostes.

La deduplicación **en destino** es la alternativa a realizarla en origen. Esta se realiza en el dispositivo de backup, por lo que se soluciona el problema de carga de trabajo en el cliente para el proceso de eliminación de datos duplicados. El problema con esta solución es que el cliente manda todos los datos sin procesarlos, por lo que aumenta la ventana de backup, requiere mayor ancho de banda y aumenta es espacio necesario para su almacenamiento, hasta que se realiza el proceso de eliminación de datos duplicados, el cual se puede hacer de inmediato (en línea), o programado (post-proceso).

La deduplicación **en línea** realiza el proceso antes de que los datos sean almacenados, reduciendo la cantidad de almacenamiento necesario para la copia, pero se penaliza el tiempo necesario para identificar y eliminar la duplicación de los datos. Por lo tanto, este método es más adecuado para un entorno con una ventana de backup de gran tamaño.

En la deduplicación **post-proceso** se almacenan los datos para posteriormente ser deduplicados cuando se desee, por lo que se necesita mayor capacidad de almacenamiento, aunque se libere el espacio al finalizar el proceso. Este método es adecuado para entornos con ventanas de backup reducidas con tiempos más estrictos.

2.2.6 Backup en entornos virtualizados

En un entorno virtualizado, es obligado realizar una copia de seguridad de los datos de la máquina virtual (sistema operativo, datos de aplicación y de configuración) para evitar su pérdida o la corrupción debido a errores humanos o técnicos. Para realizar backup en este tipo de entorno hay dos posibles soluciones, **backup tradicional** y **backup basado en imágenes**, siendo muy importante el uso de técnicas de optimización, como la deduplicación, para reducir significativamente la cantidad de datos que se copian en este tipo entorno, puesto que en la mayoría de casos las máquinas virtuales comparten configuraciones similares.

En el **backup tradicional**, un agente de backup se instala en la máquina virtual (VM) o en el hipervisor. Si el agente está instalado en una máquina virtual, esta se comporta como un servidor físico para el agente, que se encarga de recopilar los datos para transferirlos al destino, pero no captura los archivos de la propia máquina virtual, como el archivo virtual de BIOS, archivo de intercambio de VM, registros o archivos de configuración. Por lo tanto, para llevar a cabo un restauración, un usuario tiene que volver a crear manualmente la máquina virtual y luego restaurar los datos.

Si el agente se instala en el hipervisor, las máquinas virtuales se comportan como un conjunto de archivos de cara al agente. Por lo tanto, los archivos se pueden copiar mediante un backup del sistema de archivos desde el hipervisor, en lugar de tener un agente por VM, lo que rebaja la complejidad.

Un método de backup tradicional puede provocar sobrecarga de CPU, ya que se debe realizar cuando los recursos de los servidores están inactivos o durante un período de baja actividad en la red.

El backup basado en agente ha pasado de ser una solución ineficiente a una solución a tener muy en cuenta gracias a la evolución que ha sufrido esta tecnología. Ahora utilizan el filtrado a nivel de volumen, mejorando sustancialmente frente a las herramientas que trabajaban a nivel de sistema de fichero, puesto que ahora, en lugar de fijarse en los ficheros que han sido modificados a lo largo del tiempo, se centran a más bajo nivel en los bloques de disco, obteniendo un rendimiento mucho más alto.

Además estos nuevos agentes permiten consolidar la plataforma de backup entre servidores físicos y virtuales, reduciendo la complejidad. También se han desarrollado agentes especiales para aplicaciones críticas determinadas, como pueden ser las bases de datos.

Por otra parte, este tipo de backup no es la mejor solución posible en grandes entornos, puesto que en este tipo de entornos las herramientas de protección de datos requieren la instalación de uno o más agentes en cada servidor o VM que sea necesario proteger, por lo que en entornos grandes es más susceptible a mostrar los siguientes inconvenientes:

- Requieren tener un programa instalado en cada VM. Si una VM no tiene los agentes, queda desprotegida y vulnerable ante una pérdida de datos.
- Pueden producir conflictos con otras aplicaciones.
- Son difíciles de administrar, y es difícil ver qué VM, tiene o no, agentes instalados.
- Utilizan CPU y memoria en cada VM.

En la siguiente figura podemos ver la arquitectura de un backup tradicional, tanto con agentes a nivel de VM, como a nivel de hipervisor.

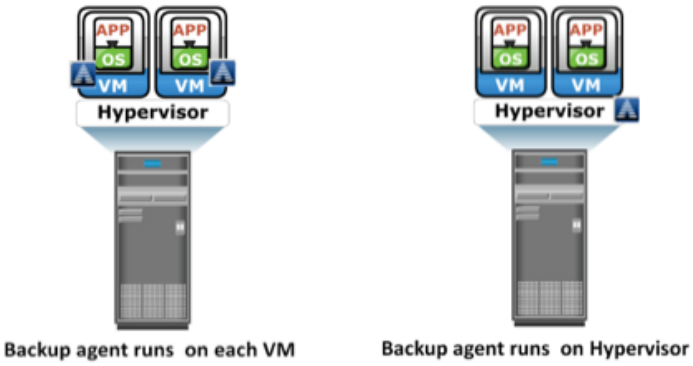

Figura 17. Backup a nivel VM e hipervisor *Copyright 2012 EMC Corporation*

En el **backup basado en imágenes** se opera a nivel de hipervisor tomando instantáneas (snapshots) de la máquina virtual, para crear una copia del sistema operativo y todos los datos asociados a ella, es decir, se toman snapshots de los archivos de las VM, incluyendo las configuraciones y aplicaciones de las VM. La copia se guarda como un único archivo llamado "imagen" y esta se monta en el servidor, actuando como un cliente de backup. El software de backup luego hace copias de seguridad de estos archivos de imagen como en el backup tradicional. Este método permite una rápida restauración de las máquinas virtuales. En la siguiente figura se muestra la arquitectura de un backup basado en imágenes.

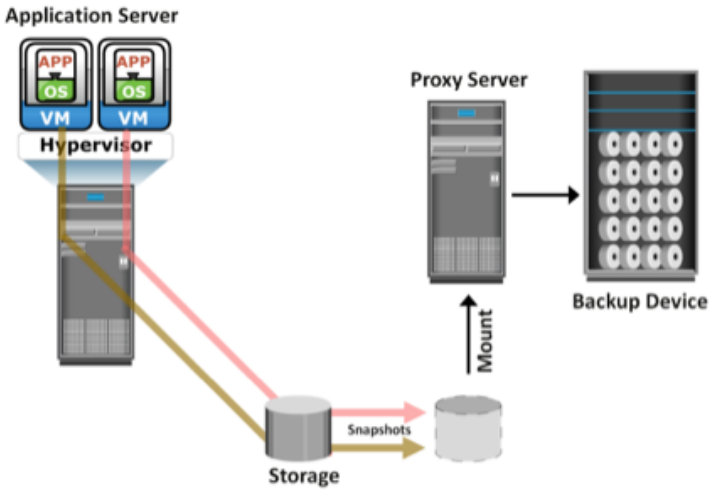

Figura 18. Backup mediante snapshot *Copyright 2012 EMC Corporation*

2.3 Replicación

Hoy en día, en los entornos empresariales, es obligado para las empresas proteger sus datos críticos, así como minimizar los riesgos para la continuidad de negocio. Cuando ocurre un desastre o los sistemas se caen provocando una interrupción del servicio, es esencial una rápida restauración y reinicio para volver a la normalidad y asegurar la continuidad de negocio. La replicación es una de las formas de conseguirlo. Se trata de el proceso de crear una copia exacta de los datos, que será utilizada para restablecer y restaurar los procesos en caso de interrupción y pérdida de datos.

La replicación se puede clasificar en dos categorías, local y remota. La primera se da cuando la réplica se realiza en el mismo centro de datos o array, mientras que la segunda, replicación remota, ocurre cuando la réplica se realiza a un centro remoto. Podemos encontrar en cada categoría diferentes tecnologías

2.3.1 Replicación local

Replicación basada en volumen lógico, el gestor de volúmenes lógicos es responsable de la creación y el control de los volúmenes lógicos a nivel de host. Un LVM tiene tres componentes: los volúmenes físicos (disco físico), los grupos de volúmenes (agrupación de volúmenes físicos) y volúmenes lógicos. Los volúmenes lógicos se crean dentro de un grupo de volumen determinado. En la replicación basada en LVM, cada bloque lógico en un volumen lógico se asigna a dos bloques físicos en dos volúmenes físicos diferentes, como se muestra en la siguiente figura. Una aplicación escrita en un volumen lógico se escribe en los dos volúmenes físicos por el controlador de dispositivo LVM, haciendo un efecto espejo o mirroring entre los volúmenes físicos.

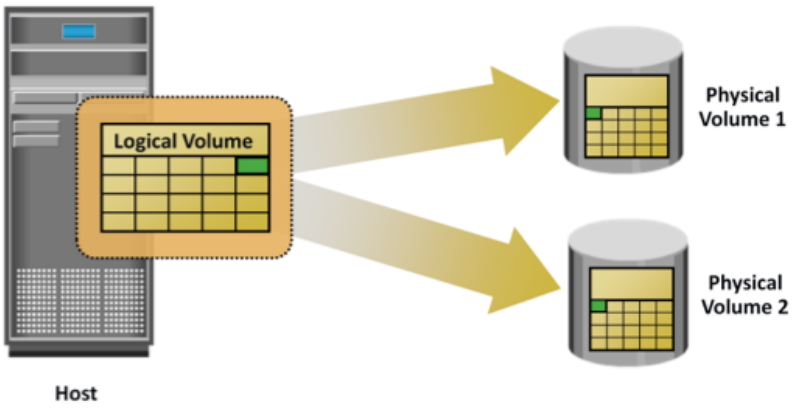

Figura 19. Replicación basada en LVM.

Este método ofrece ventajas como no depender de un único fabricante del sistema de almacenamiento, así como no requerir de licencias adicionales al formar parte del sistema, por lo que no supone sobrecostes. Por otro lado, también tiene varias limitaciones. Cada escritura que genera una aplicación se traduce en dos escrituras en el disco, lo que provoca una sobrecarga en la CPU del host, perjudicando el rendimiento. Además, la presentación de una réplica local a otro host no es posible, ya que la réplica seguirá siendo parte del grupo de volumen.

Replicación basada en snapshot del sistema de fichero es una réplica mediante puntero que requiere parte del espacio usado por el sistema de ficheros en producción. Se utiliza la *copia en la primera escritura* (CoFW) como principio para crear los snapshots. Cuando se crea el snapshot, se genera un mapa de bit y uno de bloque en los metadatos del Snapshot FS. El mapa de bit se usa para mantener la pista de los bloques que han cambiado en el sistema de producción después de un snapshot, mientras que el mapa de bloque se usa para indicar la dirección exacta de los datos que han de ser leídos para que sean accedidos desde el sistema de snapshot. Justo después de la creación de un snapshot, todas las lecturas se sirven desde el sistema de producción.

En un mecanismo CoFW, si una escritura de E/S se realiza en el FS de producción por primera vez después de la creación de un snapshot, la E/S se lleva a cabo y los datos originales del FS de producción correspondiente a esa posición se mueven al snapshot FS. Para realizar la lectura del Snapshot FS se consulta el mapa de bit, si es 0, la lectura se sirve mediante el sistema de producción, mientras que si el valor del bit es 1, la dirección del bloque se obtiene del mapa de bloque y los datos se leerán desde esa dirección en el sistema de fichero del snapshot. En la siguiente figura podemos ver un diagrama de este tipo de replicación.

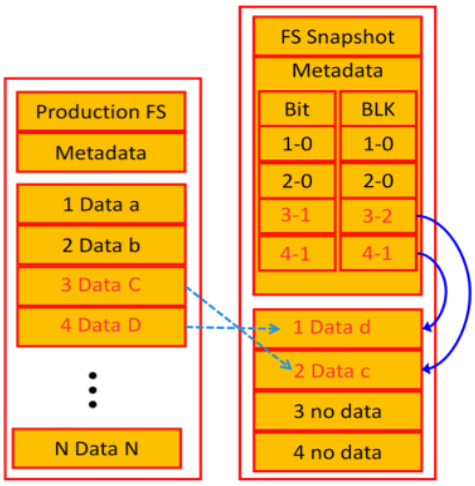

Figura 20. Replicación basada en Snapshot File System. *Copyright 2012 EMC Corporation*

Otro tipo es la replicación **basada en array de almacenamiento** donde el propio array realiza el proceso de replicación local. Con esta solución, los recursos del propio host, como la CPU y la memoria no se utilizan en el proceso de replicación. Por lo que liberamos al host de las operaciones de replicación. Además se puede acceder a la réplica mediante un host alternativo, cosa que no podíamos hacer con las anteriores soluciones basadas en host.

En esta réplica, el número requerido de los dispositivos de réplica se debe seleccionar en el mismo array para que luego los datos sean replicados entre los pares origen-réplica. En la siguiente figura se muestra un ejemplo de esta replicación local basada en array de almacenamiento, en el que el origen y el destino se encuentran en la misma matriz. Este tipo de solución se suele implementar habitualmente en tres formas: replicación mirroring de volumen completo, replicación mediante puntero de volumen completo y replicación virtual mediante puntero.

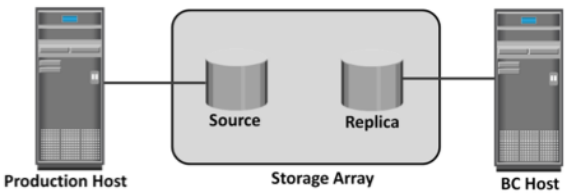

Figura 21. Replicación basada en Storage array. *Copyright 2012 EMC Corporation*

En la replicación **full-volume mirroring** el destino está unido a la fuente y se establece como un mirroring del volumen completo del origen, además las nuevas actualizaciones en origen se actualizan en el destino para mantener los datos sincronizados. Mientras sucede esto, no se puede acceder a la réplica por otro host, pero si al destino. Después de la sincronización, la réplica se separa de la fuente y se pone a disposición, para que ahora, el origen y el de destino puedan acceder a las operaciones de lectura y escritura de producción y de continuidad del negocio respectivamente. La siguiente figura muestra este proceso.

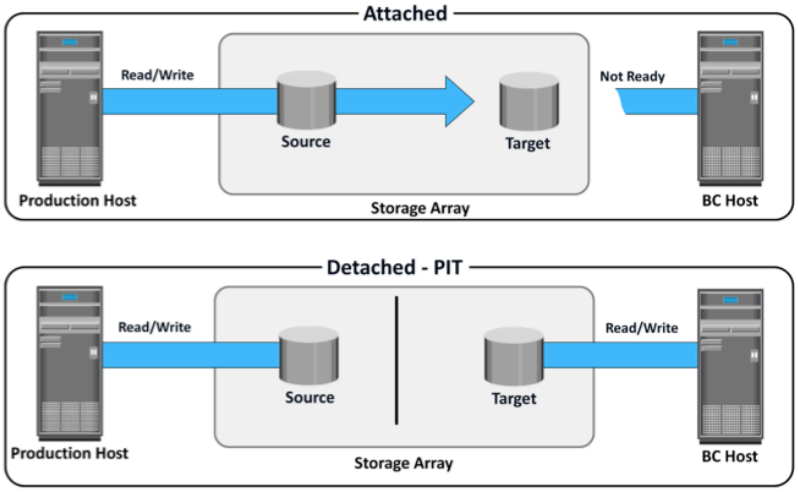

Figura 22. Full-volume mirroring.

De forma similar a la anterior, tenemos la replicación **full-volume mediante puntero** que realiza copias exactas de los datos de la fuente sobre el dispositivo de destino, con la diferencia de que con esta solución, la réplica está disponible para otros host de forma inmediata después de que la sesión de replicación este activa. Además, no requiere de sincronización y posterior desmontado para poder acceder a la réplica. Este método puede funcionar mediante CoFA o Full copy.

Por último, en **la replicación virtual mediante puntero** el destino contiene punteros que apuntan a la localización de los datos de la fuente, por lo que el destino, llamado réplica virtual, no contiene en ningún momento los datos y al igual que el método anterior, en todo momento permanece accesible después de la activación de la sesión de replicación. Esta solución usa CoFA y normalmente se usa cuando los cambios de los datos en la fuente son inferiores al 30%.

Para realizar la restauración de las réplicas full-volumen (tanto en modo mirrorin como mediante punteros) se pueden restaurar a los dispositivos de la fuente original, donde la restauración puede ser incremental o a un nuevo conjunto de dispositivos, donde esta debe ser completa. Además, para restaurar en la replicación mediante punteros, tanto virtual como total en modo CoFA, el acceso a los datos depende de la

salud y la accesibilidad de los mismos, si la fuente es innacesible, las réplicas no se podrán usar para restaurar o reiniciar el sistema, al no estar disponibles los datos que son apuntados por los propios punteros que contiene la réplica, como es lógico.

Otro tipo de replicación es la **basada en red**, que se produce en la capa de red entre los hosts y los arrays de almacenamiento. La replicación basada en red combina los beneficios de la basada en host y en storage-array, puede trabajar a través de un gran número de servidores y sistemas de almacenamiento, por lo que es ideal para entornos altamente heterogéneos. En este tipo de replicación se introduce una nueva tecnología llamada protección continua de datos, CDP, del inglés **Continuous Data Protection**, que se usa para replicaciones locales y remotas basadas en red.

En un centro de datos, las aplicaciones críticas requieren puntos de recuperación de datos instantánea, mientras que en las tecnologías tradicionales estos puntos son más limitados. Con CDP, cuando se produce la pérdida de datos, el sistema puede recuperar automáticamente el último punto de recuperación disponible, mediante el seguimiento de todos los cambios en los dispositivos de producción.

CDP utiliza un volumen diario para almacenar todos los cambios de datos en el almacenamiento principal, que contiene todos los datos desde que se inició la duplicación. La cantidad de espacio que se configura para este volumen determina la cantidad de puntos de restauración. Además se usa un appliance, que es una plataforma de hardware inteligente, que se encarga de ejecutar el software CDP y gestionar la replicación de datos locales y remotos. En la siguiente figura se muestra gráficamente el funcionamiento de CPD.

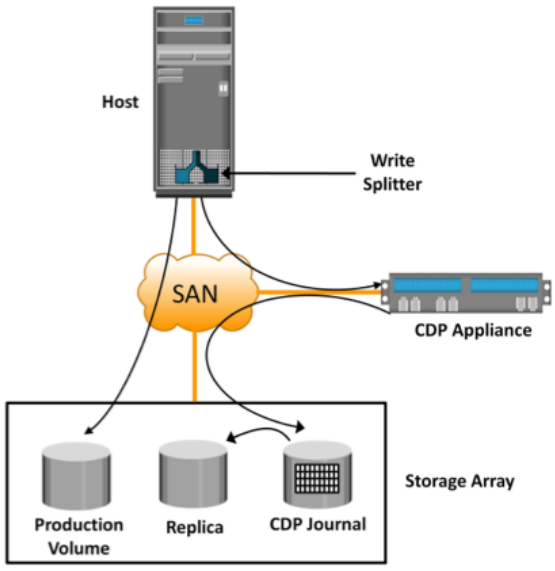

Figura 23. Continuous Data Protection.

2.3.2 Replicación remota

La replicación remota es el proceso para crear réplicas de la información en ubicaciones remotas. Este tipo de replicación ayuda a las organizaciones a mitigar los riesgos en caso de caída en origen o replicación local. De forma similar a las réplicas locales, las réplicas remotas también se pueden usar para otras operaciones. Podemos distinguir dos modos de replicación remota, la replicación sincrónica y asincrónica.

La **replicación síncrona** es el nivel más alto posible para los objetivos de punto de recuperación (RPO) y de tiempo de recuperación (RTO) en la recuperación ante desastres. Este tipo de replicación funciona de manera que no se completan ni reconocen las operaciones de escritura locales hasta que no se han completado y reconocido las operaciones de escritura remotas. Las operaciones de escritura adicionales no tendrán lugar hasta que no se hayan completado y reconocido cada una de las operaciones de escritura anteriores. Esto implica que el rendimiento local está directamente relacionado con el rendimiento del dispositivo de DR remoto; la distancia es el factor limitante, por lo que rara vez se implanta en circuitos con distancias superiores a 160 km.

Mediante la **replicación asíncrona**, las operaciones de escritura locales quedan completadas y reconocidas antes que las operaciones de escritura remotas. La replicación remota asíncrona es una técnica de (store-and-forward), que reduce las E/S y los retrasos por esperas, permitiendo que las operaciones de escritura remotas se completen tras las locales. Esto significa que el RPO por la pérdida de datos puede variar de segundos a minutos, y en algunos casos incluso horas, por lo que este tipo de replicación se utiliza con más frecuencia cuando el sitio remoto se halla a una gran distancia del sitio local.

La principal ventaja de la replicación remota tanto síncrona como asíncrona es la mínima o nula exposición al riesgo de pérdida de datos durante un desastre. Una ventaja adicional es la posibilidad de una rápida recuperación de datos cuando se produce un desastre. La replicación remota no requiere de agentes en el servidor, y ofrece soporte para servidores y aplicaciones heterogéneos.

Existen varias tecnologías en la replicación remota, al igual que en la local, que las podemos separar en replicación basada en host, array y en red.

En cuanto a la **replicación vía host**, se destacan dos tipos, basada en volumen lógico (LVM) y basada en envío de log. La replicación vía LVM funciona de forma similar a la replicación local, solo que el LVM se replica al host remoto, pudiendo realizarse de forma síncrona o asíncrona. En cuanto a la replicación vía log shipping, se suele usar en entornos de bases de datos. Primero se replican los componentes relevantes de la base de datos al sitio remoto, cuando el origen no se encuentra en producción. Una vez en producción, las transacciones desde el origen de la base de datos se capturan en forma de log y se transmiten de forma periódica al host remoto, que a su vez, aplica estos logs sobre la base de datos remota que fue replicada anteriormente, y que se encuentra en standby.

La replicación remota **basada en array de almacenamiento** se divide en tres métodos: síncrona, asíncrona y buffer de disco. En cuanto a las dos primeras, se comportan de forma similar a la replicación local, cambiando las características de replicación según el tipo de las misma, síncrona o asíncrona, como se ha explicado anteriormente en cada tipo. La replicación en modo buffer de disco consiste en una combinación de tecnologías de replicación local y remota. Primero se realiza una copia consistente al dispositivo local desde el host de producción, mientras el enlace al sitio remoto permanece desactivado. Después de la primera copia, se realiza una réplica local sobre ese mismo array origen. Una vez realizada, se establece conexión con el array remoto para realizar la réplica desde el dispositivo de réplica origen al destino. Una vez replicado, se vuelve a suspender la conexión entre los sites para que se produzca la réplica local en destino. En la siguiente figura se muestra un diagrama del proceso.

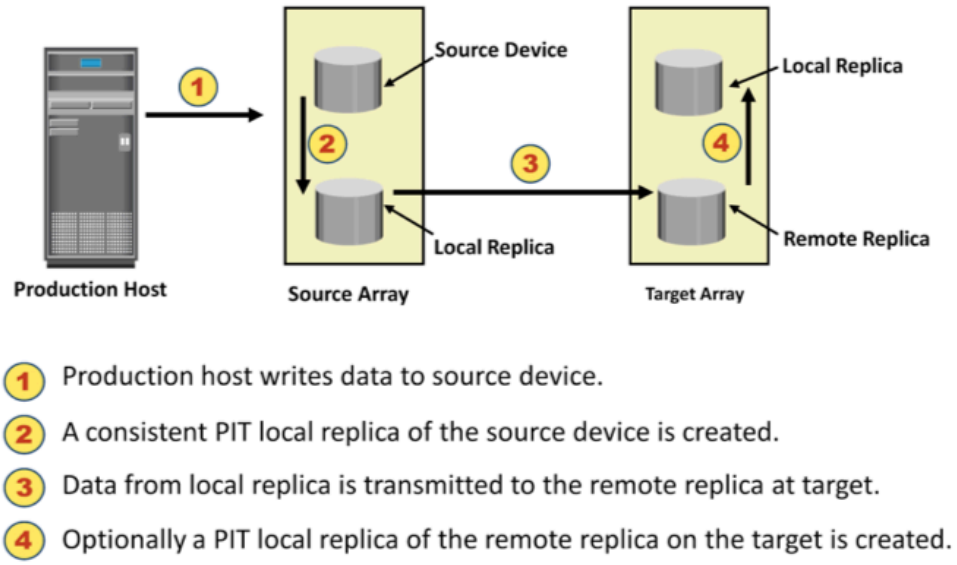

Figura 24. Proceso de backup en storage array. *Copyright 2012 EMC Corporation*

Este tipo de replicación, comparada con la síncrona y asíncrona, requiere menor ancho de banda para el proceso, debido a que la sincronización entre fuente y destino pueden ser incrementales, lo que mejora el rendimiento y los tiempos de réplicado.

Por último tenemos la **replicación remota vía red** donde la replicación ocurre a nivel de red entre el host y el dispositivo de almacenamiento. En este tipo de réplica remota también se usa la protección continua de datos, CDP, explicada en la replicación local. De forma similar a la local, también una journal volumen, appliance o software en origen y destino, así como write splitter, para llevar a cabo la replicación entre sites. En la siguiente figura se muestra el proceso de replicación remoto mediante CDP, así como los componentes esenciales.

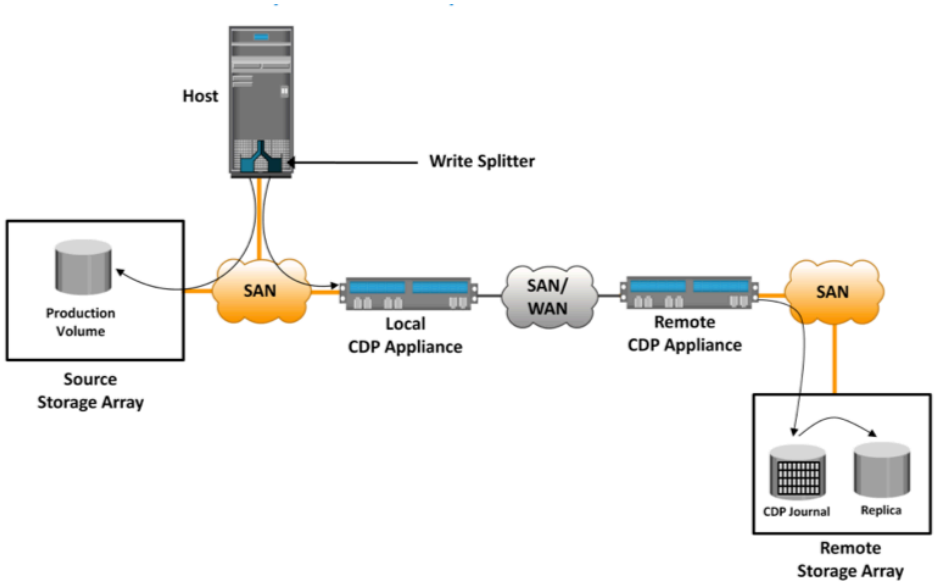

Figura 25. Remote Continuous Data Protection.

Primero la replica se sincroniza con la fuente, lo que da comienzo el proceso. Un vez empieza, se realizan dos copias, mediante el splitter, una al volumen de producción y otra al appliance local, todo mediante SAN. Después, el appliance local replica hacia el appliance destino mediante SAN/WAN

3 ESTUDIO DE MERCADO

Hoy en día existen multitud de alternativas para dar solución a los problemas de contingencia que tienen las empresas, tanto en sistemas de copias de seguridad como en sistemas de recuperación ante desastres. Además, las empresas se ven obligadas a invertir en este tipo de soluciones cada día más, debido a la cantidad de datos que mueven, que se va incrementado día a día muy rápidamente.

El problema radica en encontrar una solución que no conlleve una inversión inicial muy elevada para la implantación de la infraestructura, tanto en los equipos necesarios a nivel de hardware como a nivel de software. Además esta infraestructura debe ofrecer un alto nivel de escalabilidad, que permita crecer en cuanto a capacidad y rendimiento sin verse alterados los componentes claves del sistema.

Una parte muy importante de la solución es que sea **multi-tenant [2]**, es decir, que en la misma infraestructura los diferentes clientes compartan recursos. Esto es clave en una plataforma de estas características para poder crear un servicio en la nube y poder ofrecerlo a nuestros clientes a un precio atractivo, manteniendo siempre nuestros costes a raya. Con un sistema de este tipo, todos los clientes de la plataforma comparten eficientemente todos los recursos disponibles a nivel de hardware, gracias al software, balanceando la carga de trabajo entre los equipos disponibles, siendo totalmente transparente para el usuario.

Por lo tanto, al no tener que desplegar una infraestructura hardware o separación física para cada cliente, esto permite reducir notablemente la complejidad y la gestión del sistema, consolidándolo en una única infraestructura para todos los clientes, lo que reduce drásticamente los costes. Por ejemplo, a la hora de actualizar alguna parte del sistema, instalar un parche del fabricante, realizar snapshots o hacer copias de seguridad locales, al ser físicamente una sola máquina que alberga los datos de varios clientes, se reduce la complejidad notablemente, haciendo dichas operaciones una sola vez y no por cada cliente.

Este tipo de plataforma multi-tenant también tiene implicaciones para los clientes. Como se ha mencionado, es totalmente transparente para el usuario, por lo que a sus ojos es como si el sistema estuviera dedicado íntegramente a él. Además, como es lógico, los datos de cada cliente están protegidos y cifrados respecto a otro cliente de la plataforma. Por lo tanto, el único impacto que tiene un sistema multi-tenant para el cliente es la reducción de costes que implica, en comparación con una infraestructura propia, permitiendo incluso prescindir de personal propio, como administradores de sistemas.

Una vez conocemos el fin de la plataforma, los componentes clave que debe tener y el público objetivo, podemos definir las soluciones existentes a día de hoy que más se acercan a la solución final que buscamos, haciendo un estudio de mercado separando la solución de backup de la de disaster recovery.

3.1 Soluciones actuales en Backup.

En el mercado actual existen multitud de soluciones de diversos fabricantes para nuestras soluciones de copia de seguridad, lo que provoca una gran dificultad para elegir uno de ellos e implementar la solución óptima en cada cliente. Por lo que en la gran mayoría de los casos, prima la reducción de costes frente a la mejor solución posible, aunque en muchos casos la mejor solución para el cliente no es la más cara. Por ello, para elegir la mejor solución se ha de estudiar la situación de cada cliente, implementado la mejor de ellas en cada situación.

En nuestro caso, conociendo el cliente objetivo al que va dirigido nuestro servicio, se ha hecho un estudio de las principales soluciones que mejor se adaptarían a ese tipo de compañías, desde la que no requiere ningún tipo de infraestructura en el cliente, hasta la que necesita de un appliance en origen para replicar contra nuestra infraestructura.

Las premisas al elegir una solución son las siguientes:

- **Adaptable a clientes heterogéneos**. Esto quiere decir que la solución tiene que servir para el cliente que ya tiene una solución de copia de seguridad y para el que tiene que montarla desde cero.
- **Alta escalabilidad**. La solución tiene que partir de un entorno pequeño, para poder afrontar los costes, y crecer según se demande el servicio por los clientes.
- **Alto rendimiento**. Se requiere un rendimiento acorde al número de clientes.
- **Flexibilidad**. La solución debe de adaptarse a las diferentes casuísticas de los clientes, adoptando soluciones físicas o virtuales y pudiendo operar con diferentes software de copias de seguridad.

3.1.1 Tecnologías analizadas

3.1.1.1 EMC Avamar

EMC Avamar es un software de respaldo y recuperación con funcionalidad integrada de deduplicación de datos que resulta ideal para proteger ambientes virtuales, sistemas NAS, oficinas remotas, sistemas de escritorio, y aplicaciones críticas.

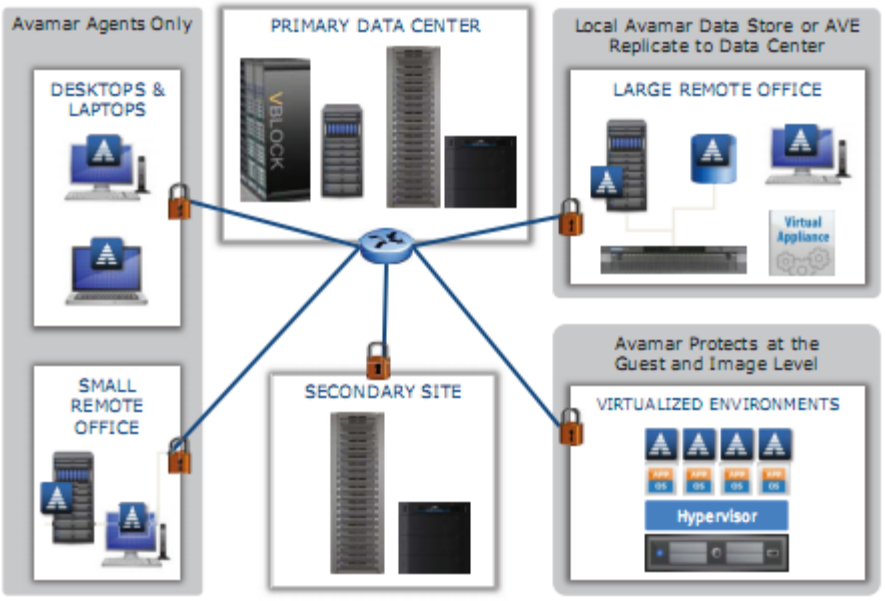

Figura 26. EMC Avamar .

El punto fundamental para no elegir EMC Avamar para la solución es debido a que su interoperabilidad es limitada, no pudiendo trabajar con software de Backup heterogéneos. Es decir, tendríamos que sustituir el software de backup en los clientes por EMC Avamar, esto incrementaría los costes y ampliarían los plazos de implementación.

La ventaja principal que se obtiene con una solución EMC Avamar es que la implementación de la deduplicación en origen hace que la utilización de las líneas de comunicaciones se reduzcan.

3.1.1.2 EMC Data Domain

EMC Data Domain es un sistema de deduplicación que se puede emplear como VTL o Backup a Disco mediante NAS.

Figura 27. EMC Data Domain

Es el sistema más parecido al elegido, siendo el líder actual del mercado dentro de los dispositivos VTL y backup a disco.

La ventaja principal que presenta es su interoperabilidad con los diferentes software de backup que existen en el mercado, así como su rendimientos y escalabilidad.

La solución EMC Data Domain no ha sido la elegida ya que no disponemos de appliance virtuales dentro de su portfolio, lo que supone un gran hándicap a la hora de desplegar el servicio en clientes con una infraestructura virtual, como se explicará más adelante.

3.1.1.3 HP StoreOnce

HP StoreOnce es un sistema de deduplicación que permite tanto VTL, NAS y se distribuye en sistemas físicos o virtuales.

Figura 28. HP StoreOnce

La ventaja principal es su interacción con HP Data Protector (Software de backup del fabricante HP) obteniendo buenos rendimientos en estos entornos, pero carece de rendimiento similar y soporte para gran cantidad de software de backup, por lo que supone un hándicap para reutilizar infraestructura del cliente, como el propio software de backup, lo que incrementaría los costes.

Su interoperabilidad con el resto de entornos es óptima pero no alcanza los rendimientos de la solución escogida de Quantum.

3.2 Soluciones actuales en Disaster Recovery

Los desastres son una realidad que todas las empresas tienen que hacer frente y deben tener en cuenta. Ya sean las fuerzas de la naturaleza, los actos de terrorismo, incendios, robos, errores humanos o acciones maliciosas, pueden causar tiempo de inactividad que perjudica la actividad de las empresas paralizando su actividad durante un periodo de tiempo determinado, lo que puede conllevar grandes pérdidas de información y económicas. Consecuentemente, la gran mayoría de empresas dedican gran parte de sus recursos a soluciones de recuperación ante desastres para garantizar la continuidad de negocio.

Muchas organizaciones hoy en día no tienen una adecuada recuperación de desastres (DR) para la protección de sus aplicaciones. En la mayoría de los casos, las soluciones recuperación ante desastres se consideran demasiada costosas, complejas y poco fiables para la mayoría de las aplicaciones de misión crítica.

La recuperación de desastres es una forma de seguro para proteger sus activos de TI cuando ocurre un desastre. Al igual que un buen seguro, la mejor recuperación de desastres debe proporcionar una gran protección, con un mínimo de molestias, al menor costo posible. VMware proporciona la protección más fiable de desastres rentable y simplificada para todas las aplicaciones virtualizadas.

Por último, en el caso del despliegue de una solución de disaster recovery, al ser aún más compleja que el servicio de backup, se llevará a cabo en una fase posterior del proyecto, una vez que se implante la solución de backup y se amortice la inversión inicial. Además, al incrementarse la complejidad y por lo tanto, el tiempo invertido para desarrollar una solución de tales características, se ha decidido post-poner esta fase de la solución, dejándola fuera del desarrollo de este proyecto y por consiguiente, de este documento.

4 DESARROLLO

4.1 Solución propuesta

El objetivo final del proyecto es la implantación de la infraestructura del centro de datos que contendrá todos los elementos necesarios para desplegar el servicio propuesto, para posteriormente ofertarlo al cliente objetivo explicado anteriormente.

Por tanto, a priori no se conoce la situación inicial del cliente, en cuanto a infraestructura se refiere, pero sí conocemos las características que tendrán este tipo de empresas, por lo que podemos hacer una previsión de las posibles soluciones que implantaremos en origen, de acuerdo a la infraestructura principal.

Por estos motivos, y teniendo siempre en cuenta el cliente objetivo, se ha optado por el fabricante Quantum como hardware principal de la solución final, debido a las grandes ventajas que ofrece, que se detallarán a lo largo de esta sección, respecto soluciones de otros fabricantes, como la facilidad de despliegue, escalabilidad, implantación en sistemas heterogéneos, sistema multi-tenant con software de monitorización para el control de costes y facturación, además de una inversión inicial con un coste contenido.

La siguiente figura muestra un diagrama general de lo que será la solución propuesta.

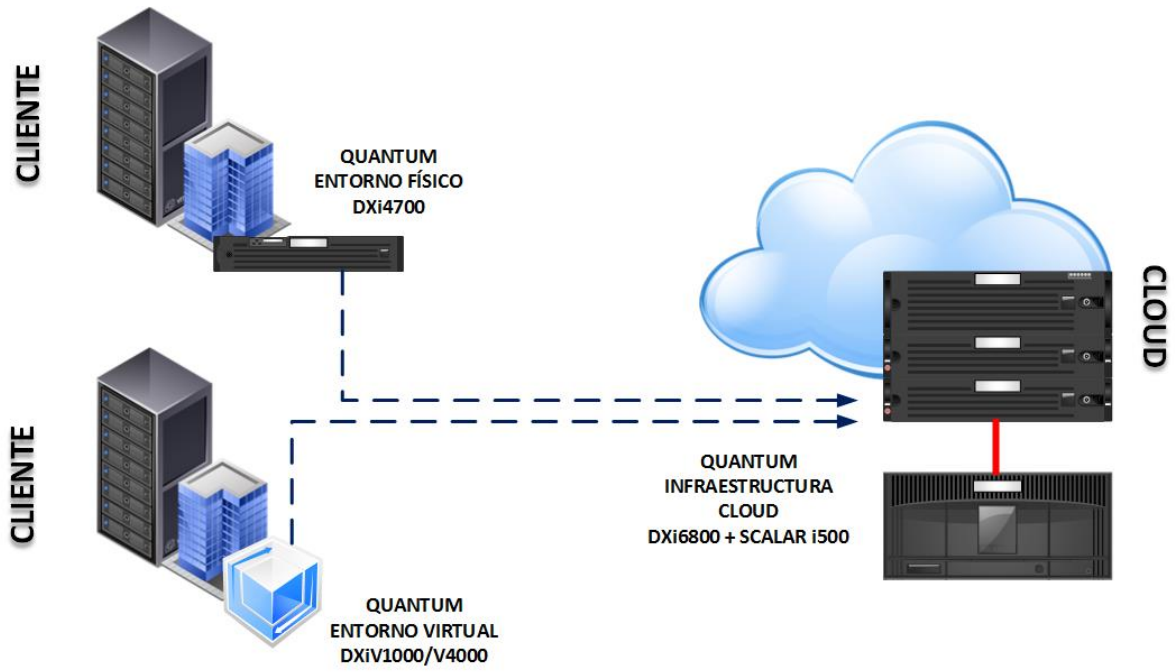

Figura 29. Diagrama general de la solución.

4.1.1 Situación inicial

En el entorno inicial de cada cliente, podemos distinguir dos posibles grupos, entornos virtuales y entornos físicos, que nos permite tener una visión más clara del tipo de infraestructura que puede tener cada empresa que requiera de esta solución, por lo que es vital para diseñar y escoger el entorno de nuestra nube. Por ello, se ha escogido la tecnología de Quantum. Además, esta solución nos permite realizar backup en sistemas heterogéneos, siendo uno de los puntos clave al tratar con una gran variedad de clientes con diferentes infraestructuras.

Por tanto, otro de los objetivos primordiales es provocar el menor impacto posible en la infraestructura ya desplegada del cliente, respetando su software de backup, en caso de disponer de él, u ofertando la mejor solución posible en otro caso. Todo ello para mantener los costes al mínimo, así como desplegar la mejor solución posible en cuanto a rendimiento se refiere.

4.1.1.1 Entornos virtuales

Para este tipo de entornos se ha optado por la serie virtual Quantum DXi, que consiste en un appliance virtual para backup con tecnología de deduplicación, que permite proteger datos físicos y virtuales a través de entornos remotos, entre el centro de datos origen (cliente) y nuestra nube. Ofrece las siguientes características:

- Elimina la necesidad de desplegar hardware adicional sin perjudicar el rendimiento
- Ideal para este tipo de entornos, BaaS (Backup as a Service).
- La tecnología de deduplicación reduce hasta el 90% el espacio necesario de almacenamiento.
- Fácilmente escalable, desde 1TB a 360TB* *(capacidad lógica usando el ratio estándar de deduplicación de 15:1)* de capacidad por appliance virtual, según el modelo elegido, DXi V1000 o DXi V4000. Lo que permite el despliegue en cualquier tipo de compañía, independientemente de la extensión, ofreciendo una gran flexibilidad.
- Rendimiento de hasta 4.9TB/hr
- Facilidad en el despliegue, pudiendo ser implantado en cuestión de minutos, como si de una aplicación más se tratase, frente a horas o días que requieren soluciones hardware, aumentando la complejidad.

En la siguiente imagen podemos ver un diagrama de la solución en este tipo de entornos virtualizados.

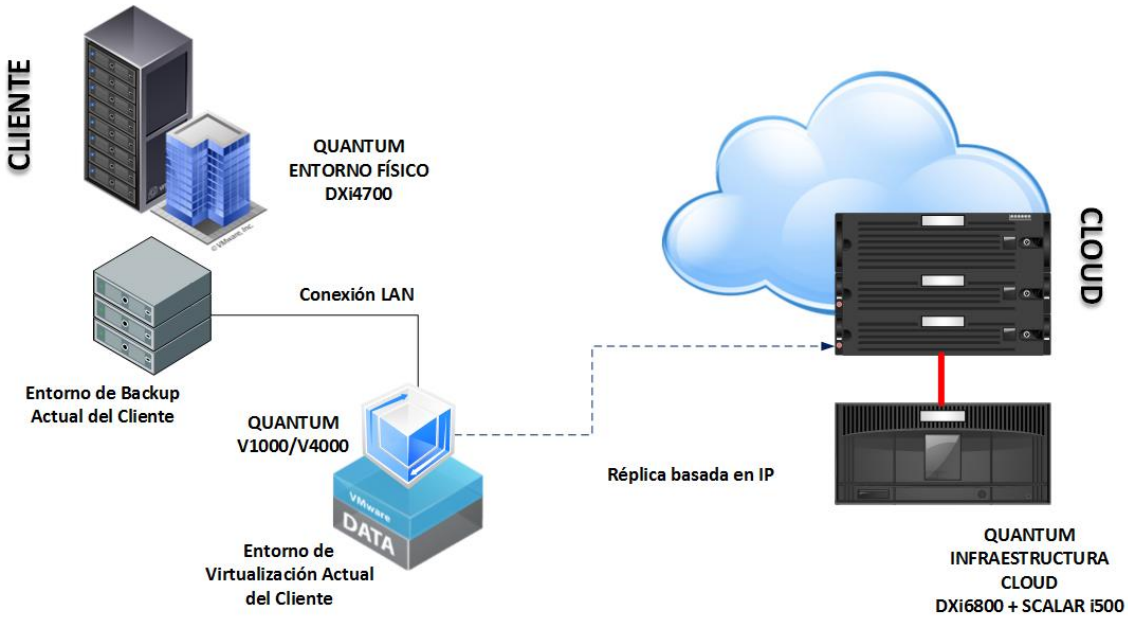

Figura 29. Diagrama general de la solución. Entorno virtual

Quantum DXi V1000 - Especificaciones técnicas

Solución ideal para la protección de datos en pequeñas empresas con capacidad de almacenamiento de hasta 30TB**(capacidad lógica usando el ratio estándar de deduplicación 15:1)* por cada appliance virtual. En la siguiente figura se muestra un esquema de la solución.

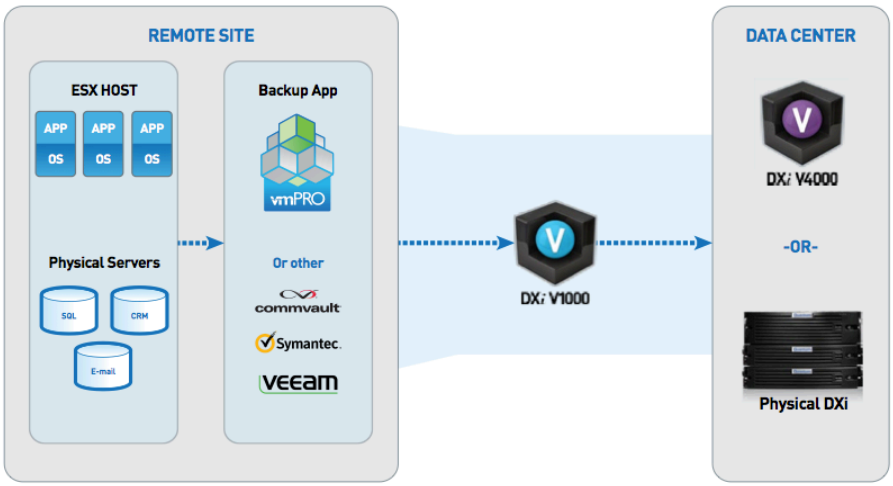

Figura 30. *Copyright Quantum Corporation*

Los datos a proteger, tanto desde dispositivos virtuales como físicos, como se puede observar en el esquema anterior, son enviados al software de backup. La aplicación de backup envía los datos recopilados al appliance virtual de Quantum, lo que proporciona una copia local para una rápida y sencilla recuperación de los datos. Los datos son replicados a nuestro centro de datos a un DXi físico, que se detallará más adelante, donde se almacenarán, según las especificaciones del cliente *(tiempos de recuperación, periodos de almacenamiento prolongado, etc.)* para completar el proceso.

Utiliza replicación asíncrona. Los datos son deduplicados y encriptados antes de la transmisión, por lo que mejoran los tiempos de la transmisión de los datos al CPD principal al transmitir menos datos, además de mejorar la seguridad en el proceso.

El software DXi Accent, incluido en toda la serie, permite al servidor de backup colaborar en el proceso de deduplicación, lo que ayuda a reducir el proceso de reducción para mandar bloques únicos de datos a través de la red hacia el appliance DXi. Esto permite copias más rápidas en ancho de banda limitados en LAN o WAN. Además, puede ser activado o desactivado según necesidades.

Características de la serie V1000:

- **Rendimiento**: Hasta 2.9TB/hr.
- **Capacidad** usable: 1TB- 2TB
- **Software incluido:** Deduplicación, replicación, encriptación, DXi Accent, NAS(CIFS/NFS), Soporte Symantec Open Storage OST.
- **Interfaces:**
  - NAS: Mediante CIFS y/o NFS. Presentación de hasta 128 objetivos.
  - OpenStorage API –OST: Mediante Symantec Storage Servers y Unidades Lógicas de Almacenamiento (LSU). Presentación de hasta 128 objetivos.

Requisitos mínimos de la serie V1000

Este modelo está pensado para empresas con entornos virtualizados con los siguientes requerimientos mínimos en sus sistemas.

- VMware ESX/ESXi 4 o 5; VMware Workstation 9.x
- VSphere 4.0 Update 2 o vSphere 5
- 4 GB RAM
- Requiere 500GB además de la siguiente capacidad dependiendo de la licencia:
  - 756GB para 256GB de licencia.
  - 1.5TB para 1TB de licencia
  - 2.5TB para 2TB de licencia
- Requiere procesadores Intel, así como 2 virtual CPU cores.

Quantum DXi V4000 – Especificaciones técnicas

Solución ideal para necesidades de alto rendimiento y capacidad, ofreciendo hasta 360TB**(capacidad lógica usando el ratio estándar de deduplicación 15:1)* por cada appliance virtual. La siguiente figura muestra un esquema de la solución propuesta.

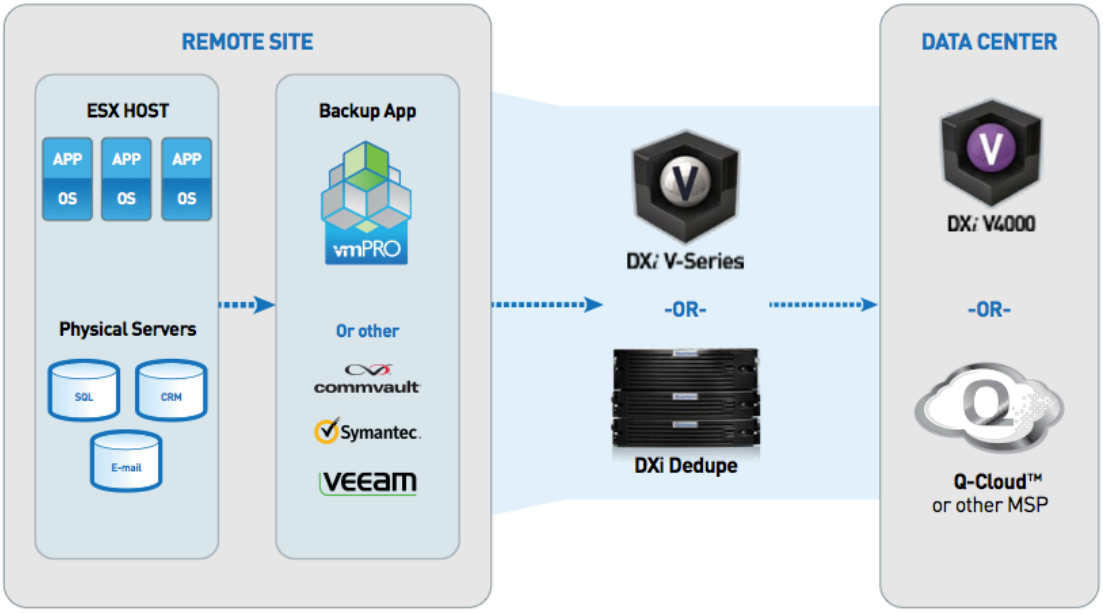

Figura 31. *Copyright Quantum Corporation*

Características de la serie V4000:

- **Rendimiento**: Hasta 4.9TB/hr.
- **Capacidad** usable: 4TB- 24TB
- **Software incluido:** Deduplicación, replicación, encriptación, DXi Accent, NAS(CIFS/NFS), Soporte Symantec Open Storage OST.
- **Interfaces:**
  - NAS: Mediante CIFS y/o NFS. Presentación de hasta 128 objetivos.
  - OpenStorage API –OST: Mediante Symantec Storage Servers y Unidades Lógicas de Almacenamiento (LSU). Presentación de hasta 128 objetivos.

Requisitos mínimos de la serie V4000

Este modelo está pensado para empresas con entornos virtualizados con los siguientes requerimientos mínimos en sus sistemas.

- VMware ESXi 5 o 5.1;
- Procesadores Intel Multi-core (AMD no soportado) y 8 vCPU.
- Mínimo de 4.5TB de capacidad de disco en modo thinly provisioning.
- Requiere 500GB además de la siguiente capacidad dependiendo de la licencia.
- 48GB de RAM dedicado al DXi V4000.

4.1.1.2 Entornos físicos

Para la pequeña y mediana empresa, donde no se dispone de un entorno virtualizado o la mayoría del mismo consiste en servidores físicos, se ha optado por la serie DXi 4700 de Quantum, que consiste en un appliance físico para backup con la mayor eficiencia en tecnología de deduplicación y replicación, además de simplicidad de escalado, que permite proteger datos físicos y virtuales a través de entornos remotos, entre el centro de datos origen (cliente) y nuestra nube. Ofrece las siguientes características:

- Ideal para este tipo de entornos, BaaS (Backup as a Service).
- La tecnología de deduplicación reduce hasta el 90% el espacio necesario de almacenamiento.
- Fácilmente escalable, desde 5TB a 135TB de capacidad por appliance desde el modelo de entrada. Lo que permite el despliegue en cualquier tipo de compañía, independientemente de la extensión, ofreciendo una gran flexibilidad.
- Rendimiento de hasta 5TB/hr
- Solución para diversos tipos de necesidades en una única plataforma, desde NAS hasta VTL.
- Conectividad desde 1GbE hasta 10GbE.

En la siguiente figura podemos ver un esquema de la solución.

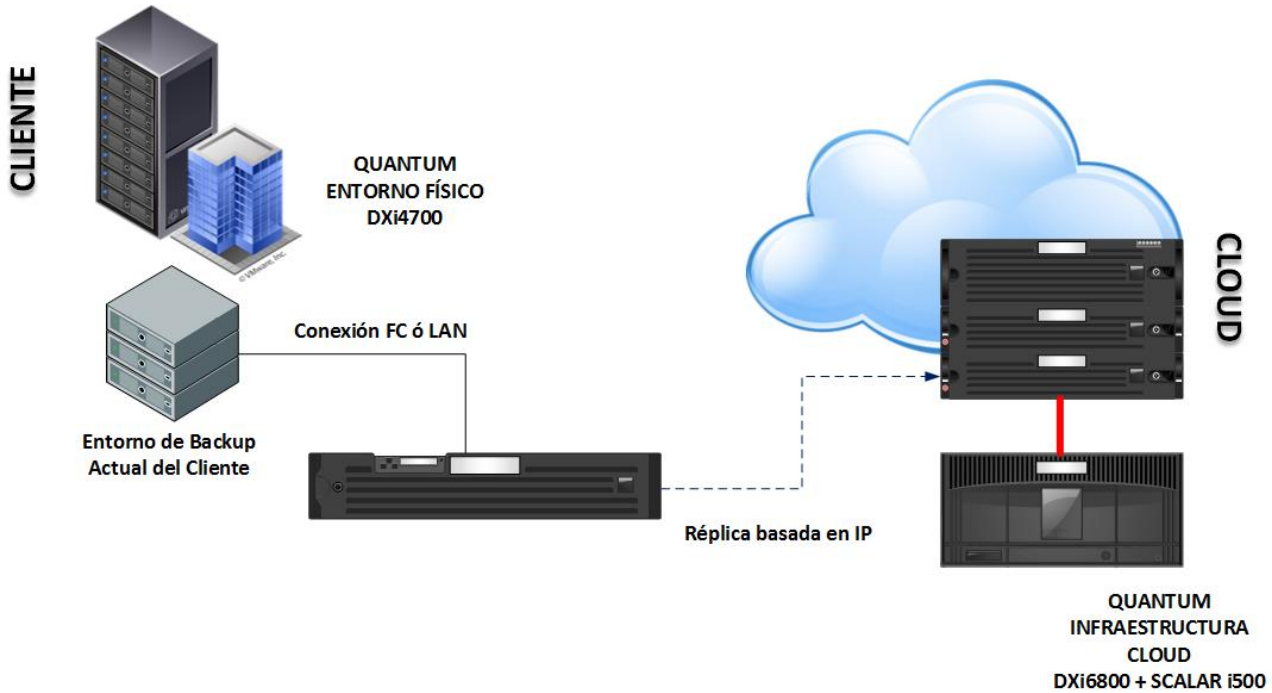

Figura 32. Diagrama de la solución. Entorno físico.

Existe otra posibilidad en este tipo de entornos con servidores físicos. Esta solución consiste en montar un servidor físico dedicado, sobre el que desplegar el mismo servicio propuesto en los entornos virtuales, mediante appliances virtuales DXi V1000 o V4000, que ofrece características similares en comparación con el appliance físico DXi 4700. La implantación de esta solución, dependerá de las necesidades del cliente.

Quantum DXi 4700

Características de la serie 4700:

- **Rendimiento**: Hasta 5TB/hr.
- **Capacidad**
  - Usable: 5TB - 135TB
  - Escalabilidad: 6TB, 8TB, o 18TB
  - Lógica: 100TB - 2700TB**(capacidad usando ratio de deduplicación 20:1)*
  - Discos: 4TB
- **Redundancia del sistema:** RAID 6, redundant power, redundant cooling, hot spare drives, hot-swap drives, power supplies y ventiladores.
- **Software incluido:** Deduplicación, replicación, encriptación, DXi Accent, NAS(CIFS/NFS), Soporte Symantec Open Storage OST.
- **Interfaces:**
  - **NAS**: Mediante CIFS y/o NFS. Presentación de hasta 128 objetivos.
  - **OpenStorage API –OST**: Mediante Symantec Storage Servers y Unidades Lógicas de Almacenamiento (LSU). Presentación de hasta 128 objetivos.
  - **VTL Fibre Channel**:
    - Particiones (max): 64
    - Drives (max): 64
    - Cintas por partición (max): 9.000
    - Emulaciones (librerías): Scalar® 24, Scalar i40/i80, Scalar 100, Scalar i500, Scalar i2000, Scalar i6000.
    - Emulaciones (drives): DLT7000, SDLT 320, SDLT 600, DLT-S4, LTO-1, LTO-2, LTO-3, LTO-4, LTO-5.
- **Especificaciones físicas:**
  - Sistema: 2U [44.5cm (W) x 8.6cm (H) x 72.6cm (D)] ; 27,7Kg
  - Modulo expansión: 2U[45.1cm (W) x 8.8cm (H) x 55.2cm (D)] ; 26,8Kg
- **Especificaciones potencia:**
  - Power Input: NEMA 5-15P a C13 power cord
  - Input Voltage: 100 a 240VAC, 50-60Hz
  - Max Power:
    - Sistema: 344W
    - Modulo expansión: 263W

4.1.2 Solución

4.1.2.1 Visión global de la solución

Una vez definido el entorno inicial que dispondrá cada cliente, podemos definir la mejor solución posible para el centro de datos principal, es decir, la infraestructura central que dará todo el servicio.

Al igual que en el entorno origen de cada cliente, dentro de las posibilidades que podemos encontrar en el mercado, se ha elegido la tecnología de Quantum, que cumple ampliamente nuestros requisitos, además de presentar grandes ventajas respecto a soluciones de otros fabricantes, como la facilidad de despliegue, escalabilidad, implantación en sistemas heterogéneos, sistema multi-tenant con software de monitorización para el control de costes y facturación, como ya se ha mencionado anteriormente.

En concreto, como hardware se ha elegido una solución basada en un sistema de almacenamiento de backups en disco Quantum DXI 6800 con deduplicación inline. Este sistema cumple ampliamente los requisitos marcados al comienzo del proyecto además de presentar ventajas importantes, postulándose como una solución ideal para entornos BaaS (Backup as a Service).

La serie DXi 6800 consiste en un appliance que ofrece una alta escalabilidad, ideal para este tipo de entornos donde se puede crecer muy rápido al depender de un número indeterminado de clientes, y que a la vez necesita de una inversión inicial mínima.

A continuación se describen las características principales del sistema mencionado:

- Sistema de almacenamiento de backups con deduplicación inline Quantum DXI Serie 6802
- Albergará backups de corta/media retención, según necesidades del cliente.
- 52 TB útiles de capacidad, con nivel protección Dynamic Disk pool
- Encriptación opcional por Hardware basada en discos SED de 3TB (Self Encryption Drives)
- Multiprotocolo:LAN (CIFS, NFS, OST) y SAN (VTL)
    - 6 x FC 8 Gbit/s
    - 3 x 1 GbE
    - 2 x 10 GbE
- Todas las opciones software incluidas (Replicación, Deduplicación, NAS, OST, VTL, DXi Accent, entre otros)
- Advanced Reporting, para monitorización detallada de los sistemas Quantum DXI.
- Quantum Vision, para reporting y monitorización centralizada de todas las soluciones Quantum del entorno, locales y remotas.

En la siguiente imagen podemos ver el aspecto del sistema elegido de Quantum, DXi6802.

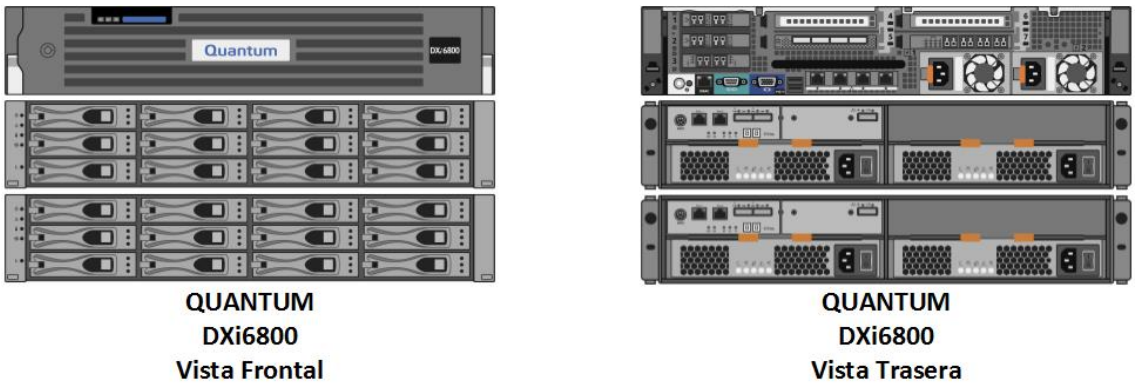

Figura 33. Frontal y Trasera Quantum DXi 6800

Este sistema puede ser configurado simultáneamente de dos formas posibles, según las necesidades del cliente:

Como VTL: Emulando una o varias librerías físicas de cintas, en concreto el modelo elegido puede emular las siguientes librerías y drives:

- Librerías: Scalar® 24, Scalar i40/i80, Scalar 100, Scalar i500, Scalar i2000, Scalar i6000.
- Drives: DLT7000, SDLT 320, SDLT 600, DLT-S4, LTO-1, LTO-2, LTO-3, LTO-4, LTO-5

Se escribirán los datos de backup en cartuchos virtuales de cintas definidos automáticamente por el sistema Quantum DXI en cada VTL creada, realizando la escritura vía SAN. Cada VTL tendrá asignados un número determinado de cartuchos virtuales en el formato que se decida (LTO3, LTO4, LTO5…).

Cada conjunto de datos de backup se guardará en uno o varios cartuchos virtuales, según el tamaño de este, y se mantendrá durante todo su período de retención asignado en el sistema Quantum DXI, excepto para aquellos backups de larga retención, donde el sistema Quantum DXI exportará de forma automática sus cartuchos virtuales a cartuchos físicos en la librería física de cintas Quantum Scalar, a través de la SAN.

Como Backup a disco (B2D): Se exportarán los puntos de montaje CIFS/NFS desde el sistema Quantum DXI al servidor de backup del cliente.

Cada punto de montaje es accesible desde los servidores o software de backup del cliente a través de dispositivos lógicos Advanced File Type Device (AFTD) como un B2D tradicional.

Cada AFTD recibe los datos de backup de los trabajos de copia de seguridad que tenga definidos a nivel de software de backup, y los datos de backup serán retenidos en el *AFTD* residente durante todo el período de retención que tengan asignado, excepto aquellos backups de larga retención, que una vez cumplido su período de retención en el sistema Quantum DXi, mediante unas políticas determinadas, según las necesidades de cada cliente, se definirán configuraciones para automatizar el traspaso automático de los datos de backup desde el sistema Quantum DXi 6802 a una librería de cintas, en concreto se ha optado por el modelo Quantum Scalar i500.

Por lo tanto, se accederá a la librería de cintas a través de dos formas, para la externalización de backups de larga retención, así externalizar cualquier backup en caso de necesidad de forma manual.

- Desde el sistema Quantum DXi en modo VTL, mediante la funcionalidad *path-to-tape*, a través de conexión SAN directa al backend del sistema Quantum DXI.
- Desde el sistema Quantum DXi en modo B2D, vía SAN, en el caso de que se ejecuten los procedimientos del software de backup en el origen, para los recursos AFTD asociados al B2D LAN al sistema Quantum DXI.

La librería de cintas seleccionada, Quantum Scalar i500, es totalmente compatible con los sistemas elegidos de la propia marca, además de ofrecer una gran escalabilidad con una arquitectura modular, un gran rendimiento y fiabilidad, así como facilidad en la gestión y manejo. El modelo seleccionado cuenta con:

- Dispone de 1 hasta 18 drives, que son los encargados de grabar las cintas.
- Dispone de 41 hasta 409 cintas.
- Funcionalidad Quantum Vision. Software para la monitorización y reporting de todos los sistemas.

En la siguiente imagen podemos ver el aspecto de la librería elegida de Quantum, Scalar i500.

Figura 34. Frontal de Quantum Scalar i500.

4.1.2.2 Sistemas Quantum DXi

Para proponer esta solución Quantum DXi, se ha realizado un análisis previo de las necesidades operativas de los futuros clientes objetivo, tratando puntos como el rendimiento y la escalabilidad, para así poder justificar el dimensionamiento de la solución inicial.

A continuación realizamos un análisis de los requerimientos mencionados, punto por punto, para poder justificar el dimensionamiento de solución realizado.

Rendimiento

Para poder evaluar de forma adecuada las necesidades de rendimiento de la solución Quantum DXi, se ha partido de la premisa de considerar el caso peor, en el que se debe acometer el backup completo/full de todos los entornos protegidos de cada cliente.

Se ha estimado un backup de un determinado tamaño, ya deduplicado, puesto que la deduplicación se realizará en origen siempre que sea posible. El tamaño de un backup completo por semana, según las características del cliente, podría estar en torno a 3TB por semana de backup ya deduplicado.

Además, hacemos una estimación inicial de 10 clientes, con el mismo tamaño de backup propuesto, 3TB/semana.

Determinamos una ventana de backup, puesto que la gran mayoría de los clientes realizarán sus copias de seguridad fuera de la jornada laboral, por lo que se realizarán los trabajos de backup durante la noche. Se ha estimado una ventana de backup de 8 horas.

Una vez fijados estos valores, podemos hacer una estimación de rendimiento y así compararla con el rendimiento que nos aporta la solución de Quantum. Por lo que tenemos:

- Alrededor de **3 TB/semana** por cliente.
- Una estimación de **10 clientes** iniciales.
- Un total de unos **30 TB/semana**
- Ventana de backup de **8 horas**.
- Rendimiento mínimo necesario → 30000 GB / 8 horas = **3750 GB / hora**

Por tanto, el rendimiento mínimo necesario, según los datos estimados, de acuerdo a mi experiencia y las tendencias de este tipo de compañías, es de aproximadamente:

- Rendimiento mínimo → **3,8 TB/hora**

En el caso de los sistemas elegidos del fabricante Quantum, superan ampliamente el rendimiento mínimo requerido en este tipo de entornos para cumplir la ventana de backup estimada, dejando margen suficiente para reducir, en caso necesario, la ventana de backup, así como ampliar el número de clientes sin ver comprometido el rendimiento.

En la siguiente tabla podemos hacernos una idea de los datos de rendimiento de los sistemas Quantum, en concreto el modelo elegido, DXi 6802, según el modo de funcionamiento, VTL o B2D.

Rendimiento Sistemas Quantum DXi		
Sistema	**Modo**	
	VTL	**B2D NAS(CIFS/NFS)**
DXi 6802	15 TB/h	13,5 TB/h

Tabla 3. Rendimiento Quantum DXi

Como se puede observar, superan ampliamente el rendimiento mínimo necesario, independientemente del modo de funcionamiento. Además estos datos, según las especificaciones del fabricante, se consiguen a través de FC (Fiber Channel), mientras que en nuestro caso, hay que tener en cuenta que las réplicas se realizarán vía IP, es decir a través de la red, por lo que el rendimiento se ve reducido, teniendo aún margen suficiente.

Por otro lado, todos los rendimientos mencionados son alcanzables sin requerir la instalación de ningún tipo de plugin, evitando así impactar en el rendimiento de los servidores del cliente.

Por lo tanto, el rendimiento proporcionado por las unidades Quantum DXi, en comparación con otros fabricantes, es el más destacado del mercado, doblando a sus inmediatos competidores que son los sistemas Data Domain de la compañía EMC, como se puede observar en el siguiente gráfico.

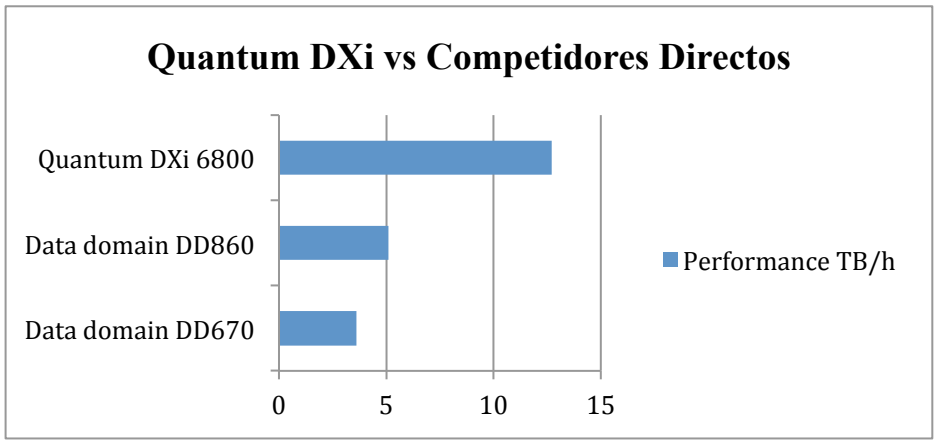

Gráfico 1. Rendimiento Quantum vs EMC DD

Escalabilidad

La solución propuesta en el caso de la opción Quantum DXi 6802, proporciona un rango de **13TB** a **156 TB útiles** (protegidos en Dynamic Disk pool) aportando una gran flexibilidad y escalabilidad. Los sistemas Quantum DXi 6802, aportan las siguientes características en cuanto a escalabilidad se refiere:

- Incrementos de 13TB útiles a través software, según licenciamiento (*capacity on demand*).
- Discos SED (self encryption drives) 3TB SATA en Dynamic Disk pool.
- Gran densidad de hasta 156TB en 14U, lo que proporciona una capacidad de hasta 11.1 TB/U.

4.1.2.3 Detalle técnico de la solución

Una vez definidos los principales componentes de la solución, pasamos a especificar el despliegue de los sistemas elegidos, así como seleccionar la localización que tendrá la infraestructura principal.

Para la localización de la infraestructura principal, se ha tenido en cuenta la cercanía a la oficina de Omega Peripherals en Madrid, para que en caso de un problema mayor, que no se pueda solucionar en remoto, el desplazamiento del técnico sea mínimo, abaratando costes.

Se ha optado por un servicio de housing, donde el proveedor nos proporcione servicios como:

- Energía eléctrica.
- Refrigeración.
- Espacio para los equipos necesarios.
- Conectividad a internet, según tipo de servicio.
- Enlaces de comunicación dentro del Data Center.

Con este servicio conseguimos de nuevo abaratar costes, al tener que invertir únicamente en los equipos necesarios para montar nuestra infraestructura, pagando una cantidad, dependiendo de los servicios contratados, por albergar nuestros equipos dentro de su centro de datos. Concretamente, se ha elegido la empresa de housing y hosting, Espacio Rack, con quién Omega Peripherals ya ha trabajado para albergar equipos de otros clientes.

Espacio Rack cuenta con un CPD situado en Madrid, concretamente en la Ciudad de la Imagen, con más de 450 metros cuadrados destinados al alojamiento de infraestructuras informáticas con altos niveles de disponibilidad y seguridad, elementos esenciales para llevar a cabo este tipo de servicios, donde la disponibilidad juega un papel esencial. Estos niveles de disponibilidad se consiguen gracias a una redundancia n+1 en todos los elementos críticos de la infraestructura.

Así mismo, cuentan con un sistema de armarios rack en cubo aislado, permitiendo así una distribución de los armarios en pasillo frío y caliente con una mayor eficiencia, donde el frío no proviene del suelo técnico, si no que se impulsa directamente al interior del pasillo frío, que se encuentra aislado, enfriando uniformemente todos los equipos, independientemente de la altura a la que se encuentren. En la siguiente imagen podemos ver la sala técnica de la Ciudad de la Imagen, donde se alojarían los equipos.

Figura 35. CPD Ciudad de la Imagen. *Copyright EspacioRack*

En cuanto a la conectividad, Espacio Rack ofrece tres tipos de velocidad de conexión hacia el exterior/internet, 10Mbps, 100Mbps y 1000Mbps. Además, se ofrecen tramos ampliables de 10Mbps, según las necesidades del cliente.

Inicialmente, se partiría de una velocidad de 100Mbps, haciendo ampliaciones de 10Mbps, según se vaya necesitando. Además, en caso de que el cliente lo requiera, se pasaría a una conexión mayor, que cubra sus necesidades.

Al tratarse de un DPaaS (Data Protection as a service), se factura mensualmente al cliente según sus necesidades, por lo que, en caso de que este requiera de algún servicio especial que no contemple nuestra infraestructura, se facturaría según sus necesidades, lo que nos permite ofrecer un servicio personalizado, manteniendo los costes a raya.

Una vez conocemos la localización de la infraestructura, podemos definir los equipos necesarios que formarán parte de ella. En la siguiente imagen podemos observar los distintos equipos necesarios para el despliegue del servicio, que irían alojados en un armario rack según se muestran.

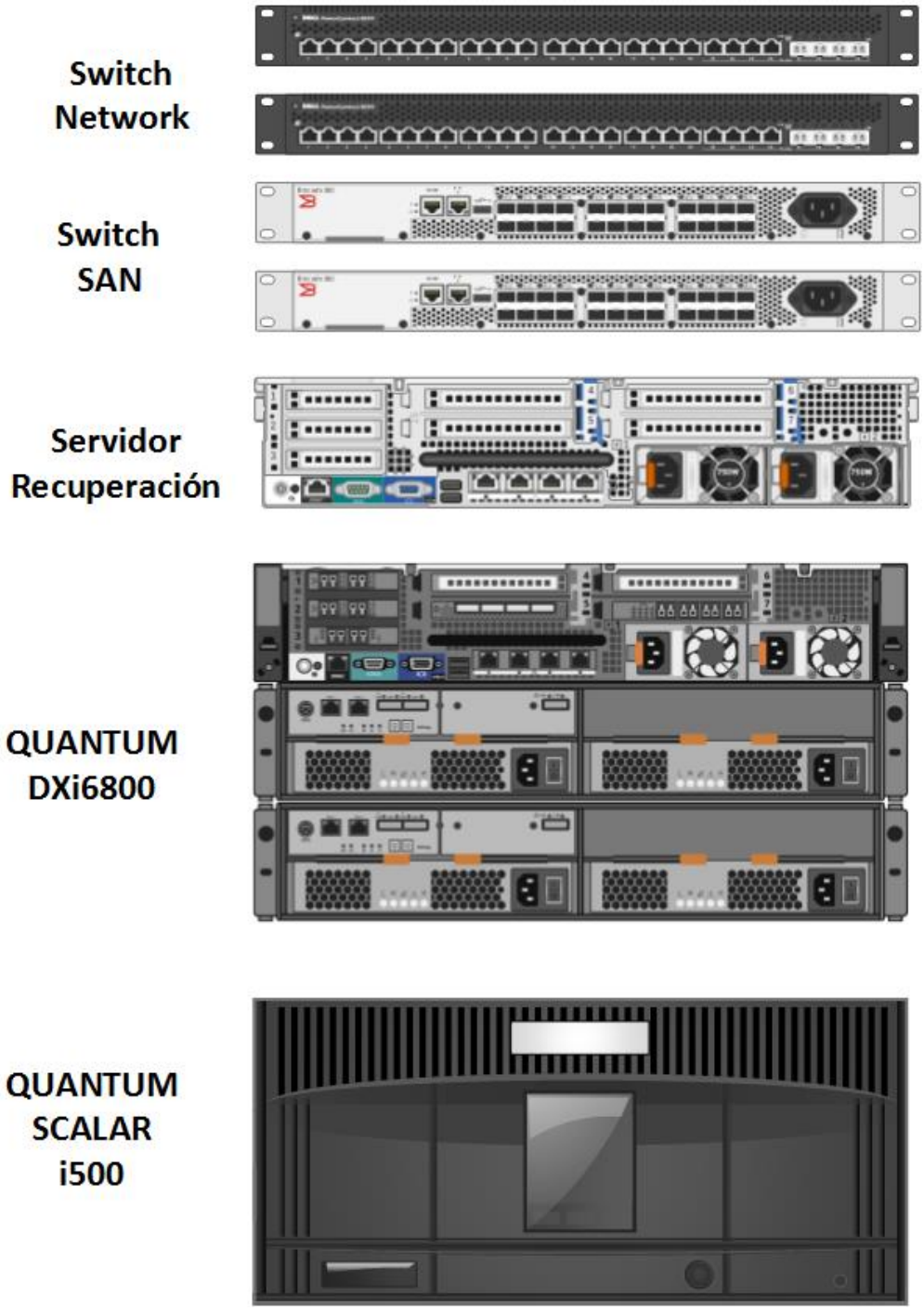

Figura 36. Diagrama general pila de equipos a instalar.

Como se puede observar en la figura anterior, comenzando desde la parte superior, en primer lugar se dispondría de dos **Switch de red**, debido a la redundancia, que permiten la conectividad LAN de los equipos, así como salida a internet, para recibir las réplicas desde los clientes.

A continuación, dispondríamos de dos **Switch de SAN,** de nuevo redundados, que conectan la red de almacenamiento.

Por último, el **servidor de recuperación,** equipo indispensable que se encargaría de seleccionar los datos necesarios en caso de que un cliente necesite recuperar los datos salvaguardados en caso de pérdida.

Los equipos comentados anteriormente, aunque indispensables para el despliegue de la infraestructura, se detallarán en la fase de implementación del proyecto, donde se elegirá la última versión del fabricante, debido al mayor número de actualizaciones que reciben este tipo de componentes, en comparación con los sistemas principales de la infraestructura, que son los sistemas Quantum DXi y Quantum Scalar.

Quantum DXi 6802

Como se ha mencionado anteriormente, es el equipo principal seleccionado para la solución. Es el epicentro de la infraestructura y dispone de las siguientes características:

Características de la serie 6802:

- **Rendimiento**:
  - **NAS:** 13.5TB/hora
  - **OST:** 16.3TB/hora
  - **VTL:** 15 TB/hora
  - **Dxi Accent:** 12.1 TB/hora
- **Capacidad**
  - Usable: 13TB - 156TB
  - Escalabilidad: 13TB
  - Lógica: 260TB - 3120TB**(capacidad usando ratio de deduplicación 20:1)*
  - Discos: 3TB SAS
- **Redundancia del sistema:** RAID 6, redundant power, redundant cooling, hot spare drives, hot-swap drives, power supplies y ventiladores.
- **Software incluido:** Deduplicación, replicación, encriptación, DXi Accent, NAS(CIFS/NFS), Soporte Symantec Open Storage OST.
- **Interfaces:**
  - **NAS**: Mediante CIFS y/o NFS. Presentación de hasta 128 objetivos.
  - **OpenStorage API –OST**: Mediante Symantec Storage Servers y Unidades Lógicas de Almacenamiento (LSU). Presentación de hasta 128 objetivos.
  - **VTL Fibre Channel**:
    - Particiones (max): 64
    - Drives (max): 256

- Cintas por partición (max): 9.000
- Emulaciones (librerías): Scalar® 24, Scalar i40/i80, Scalar 100, Scalar i500, Scalar i2000, Scalar i6000.
- Emulaciones (drives): DLT7000, SDLT 320, SDLT 600, DLT-S4, LTO-1, LTO-2, LTO-3, LTO-4, LTO-5.

- **Especificaciones físicas:**
  - Sistema: 2U [44.5cm (W) x 8.6cm (H) x 75.4cm (D)] ; 29,4Kg
  - Modulo expansión: 2U[45.2cm (W) x 8.6cm (H) x 55.4cm (D)] ; 25,9Kg
- **Conexiones:**
  - 6 x 8Gbps FC
  - Hasta 4 x 10 Gbps SFP+ Cooper
  - Hasta 7 x 1 GbE
- **Especificaciones potencia:**
  - Power Input: NEMA 5-15P a C13 power cord
  - Input Voltage: 100 a 240VAC, 50-60Hz
  - Max Power:
    - Sistema: 490W
    - Modulo expansión: 225W

Quantum Scalar i500

Características de la serie Scalar i500:

- **Capacidad**
  - 5U Modulo de control: Max- 2 drives; 41 slots
  - 14 U: Max- 6 drives; 133 slots
  - 23 U: Max- 10 drives; 225 slots
  - 32 U: Max- 14 drives; 317 slots
  - 41 U: Max- 18 drives; 409 slots
- **Operación :**
  - **Interfaz Drive:** 8Gb fibra, 4Gb fibra, 6Gb SAS, 3Gb SAS, SCSI-2/-3
  - **Interfaz Libreria:** 8Gb fibra, 4Gb fibra, 6Gb SAS, 3Gb SAS, SCSI-2/-3
  - **Velocidad de inventario:** 55 seg/5U ;110 seg/14U
  - **Configuración:** Auto descubrimiento y auto calibración para añadir nuevos módulos, cintas, drives, etc.
  - **Importación/Exportación:** Hasta 54 posiciones LTO
- **Especificaciones físicas:**
  - Sistema: 5U [44.2cm (W) x 21.9cm (H) x 79.8cm (D)] ; 30Kg
  - Sistema: 14U[44.2cm (W) x 61.9cm (H) x 79.8.cm (D)] ; 56,8Kg

4.1.2.4 Caso de uso

Supongamos que nos contacta un cliente que quiere externalizar su backup con un servicio como el nuestro, para abaratar costes de personal y reducir la inversión para tal fin. Los pasos a seguir en este supuesto caso serían los que se detallan a continuación.

Lo primero que se haría es una consultoría de la infraestructura actual del cliente, detallando los posibles puntos críticos que han de ser salvaguardados, así como todos los datos que el cliente desea que sean respaldados.

Una vez hecha la consultoría, tenemos que el cliente dispone de una infraestructura virtual, con una infraestructura desplegada de la compañía VMware, mediante su suite vSphere, que alberga todos los servidores virtualizados de la empresa. Además, el sistema actual de copia de seguridad lo compone el software de backup, EMC Networker, que se encarga de recopilar todos los datos actuales para hacer el respaldo sobre una cabina de almacenamiento, directamente a disco.

Los principales problemas que se encuentra dicho cliente, según la consultoría realizada, son de espacio de almacenamiento para sus copias de seguridad, lo que provoca que no puedan mantener un periodo mínimo de retención, además de los altos costes que supone la actualización de dicha cabina, en cuanto a hardware nuevo con mejores características que la actual, así como la compra de mayor número de discos, para ampliar el espacio disponible de la cabina. Por otra parte, disponen de licenciamiento actualizado del software de backup mencionado, EMC Networker, por lo que no sería necesario invertir nuevamente.

Una vez conocemos la infraestructura, podemos elegir una solución de las propuestas anteriormente.

En este caso, al disponer de todo su entorno virtualizado, nos decantaríamos por la solución de entornos virtualizados, descrita anteriormente. Concretamente, al no necesitar de gran capacidad y rendimiento, la solución más óptima sería mediante el sistema Quantum Dxi V1000, que al ser software, se instalaría sobre la plataforma virtualizada del cliente, aprovechando toda la infraestructura actual.

Al disponer de software de backup, y ser éste totalmente compatible con la solución propuesta, se aprovecharía reduciendo los costes del despliegue. Además, la cabina actual se mantendría, siendo suficiente para realizar el backup semanal, y en caso de pérdida la recuperación sea más rápida al no tener que hacer una recuperación remota. También se dispondría de mayor capacidad en dicha cabina, al realizar deduplicación de los datos.

Por último, el sistema V1000 replicaría todos los datos de backup, ya deduplicados, a la infraestructura central, donde se almacenarán con la retención deseada por el cliente.

En el siguiente diagrama se puede ver un ejemplo de lo que podría ser la solución final en este supuesto cliente.

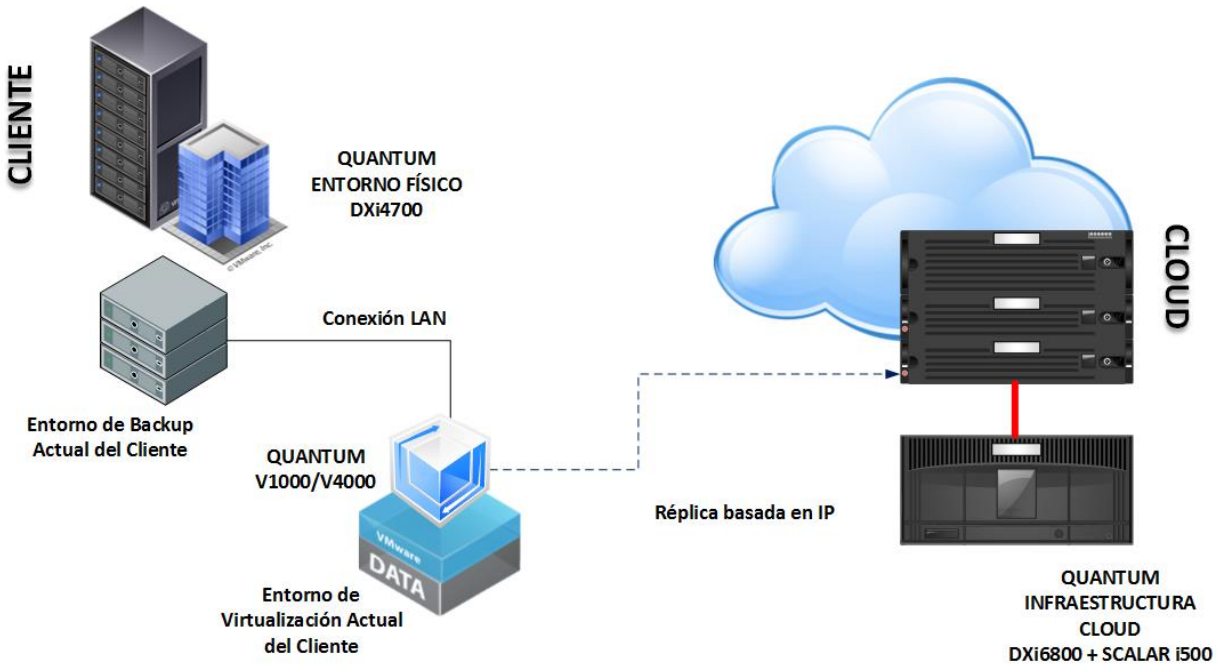

Figura 37. Diagrama general de la solución. Entorno virtual

Para realizar la facturación al supuesto cliente, esta se llevaría a cabo mediante la cantidad de datos, medidos por GiB, de los que se desea realizar el backup, además del periodo de retención que se desee, aplicando un multiplicador según los meses de retención elegidos. A continuación podemos ver los diferentes baremos a aplicar.

Multiplicador según meses de retención

1 GiB * 1 = 1 mes de retención

1 GiB * 1.2 = 2 meses de retención

1 GiB * 1.3 = 3 meses de retención

1 GiB * 1.4 = 4 meses de retención

1 GiB * 1.x = x meses de retención

Cualquier servicio adicional que se requiera se facturará a parte, como la velocidad de conexión, que correrá a cargo del cliente en caso de requerir una velocidad mayor de 100 Mbps, facturando por tramos de 10 Mbps, como se ha comentado anteriormente.

Además, se incluyen 2 horas de servicio técnico, así como 5 al mes en caso de pérdida.

Una vez consumidas dichas horas, se facturará según el tiempo de respuesta, siendo 8x5, 13x5 y 24x7, según la siguiente tabla:

Horario	Tipo de intervención	Intervención mínima	Tramo de consumo
Laboral (9:00 -19:00)	Presencial	2 horas	30 minutos
	Remoto	½ hora	15 minutos
Fuera de horario laboral	Presencial	4 horas	1.5 horas por cada hora de trabajo real
	Remoto	1 hora	1.5 horas por cada hora de trabajo real

Tabla 4. Consumo de horas para facturación.

4.1.3 Ventajas de la solución propuesta

La solución propuesta permite y ofrece las siguientes ventajas:

- **Definir una plataforma que ofrezca a los usuarios el mejor acceso posible a la datos de backup durante todo su ciclo de vida (corta retención → Disco/DXI, larga retención →Cinta/Scalar)**, tanto en términos de rendimiento como de flexibilidad, minimizando siempre los costes operacionales de la infraestructura propuesta.

- **Conseguir flexibilidad y escalabilidad** en los sistemas de almacenamiento de backups que intervienen en la solución, con el objeto de que la tecnología a implantar, sea capaz de absorber los cambios y ampliaciones futuras con el menor impacto y coste posible.

- **Conseguir un nivel de servicio óptimo** para todo el entorno de backup, mediante la utilización de las herramientas hardware y software (GUI DXI, Advanced Reporting, Vision…) que mejor se adecúan a dicho escenario.

- **Facilitar la administración del entorno**, mediante recursos hardware y software que permiten el control y la administración de los elementos ofertados de una manera sencilla, automatizada y fiable (CLI DXI, GUI DXI, Advanced Reporting y Vision, entre otros), además de la monitorización y control de costes por cliente, con la herramienta Vision.

- **Reducir al máximo los costes de inversión, mantenimiento y operación** de la infraestructura de backup, mediante la deduplicación de datos de backup, y replicación asíncrona de la información con ajuste de ancho de banda utilizado.

- **Conseguir el mejor grado de integración posible con las actuales herramientas de protección de datos**, que proporcione un rendimiento óptimo, nunca limitado por el mismo software base del sistema y que permita una total confianza en la capacidad de restauración de cualquier información perdida y respaldada por dichas funcionalidades software. Permitir de forma nativa el mejor rendimiento, evitando así el impacto en rendimiento en servidores de backup o servidores productivos.

- **Crear las bases para un esquema de contingencia del entorno de backup**, que garantice la protección y consistencia de los datos, y que asegure los menores RTO y mayores RPO (replicación asícrona IP de datos deduplicados, ajustable al ancho de banda disponible, programable y capaz de ser bidireccional).

- **Asegurar los menores costes totales de inversión**, en función de los RTOs y RPOs alcanzados, utilizando todas las funcionalidades que nos ofrecen los sistemas de almacenamiento propuestos (RAID-6, compresión, deduplicación, exportación automática a cinta física /path-to-tape, réplica remota vía IP datos deduplicados, etc.).

A continuación, se resumen los valores diferenciales de la solución Quantum DXI de la solución propuesta, que ofrece el mejor rendimiento del mercado, así como la mayor flexibilidad para adaptar la solución a cada cliente.

- Algoritmo de deduplicación inline patentado (DXI Ver. 2.0), estándar del mercado y utilizado bajo licencia por la competencia, totalmente diseñado y optimizado para deduplicación inteligente de datos de backup, con capacidad de autodetección de metadatos de software de backup utilizado, así como de tipo de datos copiados.

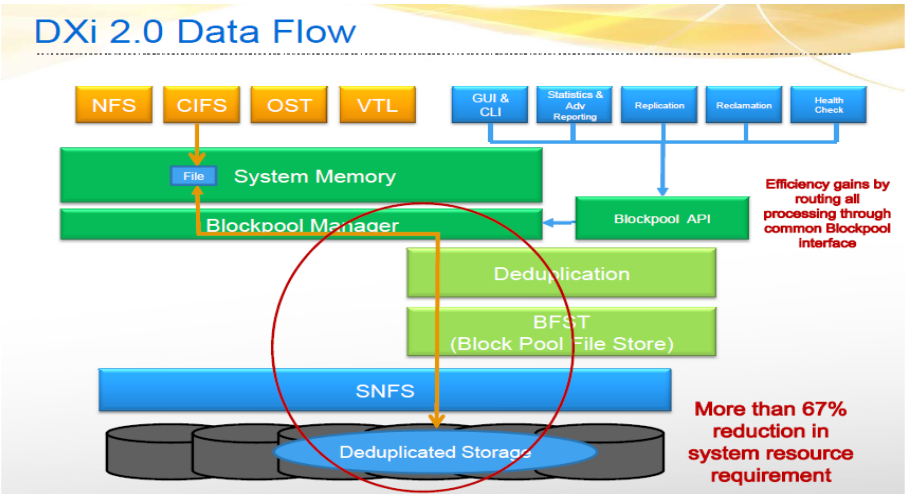

Figura 38. Quantum DXi 2.0. Flujo de datos. Quantum Copyright

- La utilización de discos de 15k revoluciones para almacenar índices de deduplicación: esta arquitectura exclusiva, combinada con el algoritmo de deduplicación, ya justifica por sí sólo el que el rendimiento sea de media un 40% superior a cualquier competidor; por otra parte, se evitan desplegar plugins intrusivos, a nivel rendimiento, en servidores de backup y storage nodes.

- La posibilidad de particiones lógicas, para cada tipo de aplicación o entorno que utilicen los sistemas Quantum DXI, con deduplicación global.

- Varias posibilidades de replicación, flexible y ajustable según el cliente, que permite modificar ancho de banda y planificación.

    - *Estándar*: toda la partición virtual (VTL o NAS) se replica a la segunda cabina. Se pueden guardar hasta 10 snapshots, para poder volver a configuraciones anteriores.

    - *Trigger- based*
        - Fichero/Directorio → replicación en caso de B2D, a nivel fichero o directorio.
        - Cinta → replicación a nivel cinta virtual, en caso de VTL

    - *OST* → Optimized (en caso entornos NBU/Backup Exec)

- Mediante la funcionalidad Advanced Reporting, incorpora las funcionalidades gráficas de reporting más completas del mercado, con todo nivel de detalle, frente a otras soluciones, como EMC Data Domain, que requieren costosas herramientas externas para proporcionar información detallada que, en cualquier caso, es más escasa que la proporcionada por Advanced Reporting (Análisis de tendencias, Performance Tuning, etc…).

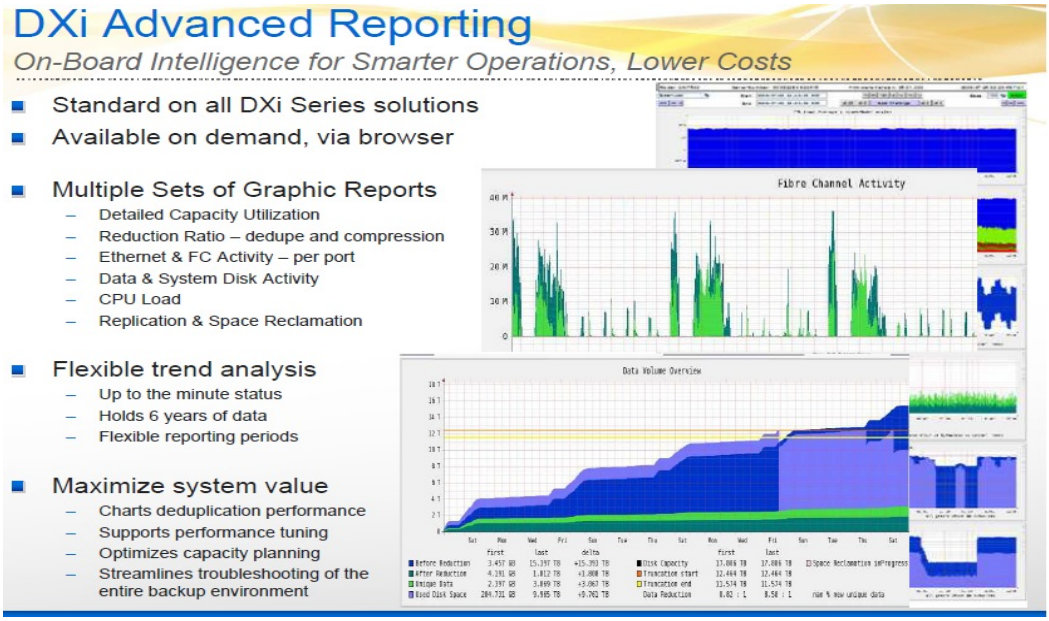

Figura 39. DXi Advanced Reporting. Quantum Copyright

- Mediante la funcionalidad Quantum Vision, la posibilidad de una monitorización y reporting unificados y centralizados de todos los sistemas Quantum DXI locales y remotos, además de las librerías Quantum Scalar (frente a soluciones individuales, no integradas de la competencia, que requieren herramientas adicionales, como EMC Data Protection Adviser).

5 CONCLUSIONES

Para finalizar la memoria de este proyecto se exponen a continuación las conclusiones a las que se ha llegado tras la realización del mismo, así como una serie de líneas futuras por las que se considera se puede seguir trabajando para contribuir en la evolución del servicio desarrollado.

El objetivo que se marcó al inicio de este proyecto, fue el desarrollo de una plataforma de servicios gestionados para soluciones de contingencia, backup y disaster recovery, en la nube. Durante la realización del proyecto se han llevado a cabo todos los objetivos marcados, relacionados con la parte de backup.

En cuanto al desarrollo de la infraestructura de recuperación ante desastres, al ser aún más compleja y necesitar de mayor cantidad de tiempo que la solución de backup, se tomó la decisión de llevarla a cabo en una fase posterior del proyecto, una vez que se implante la solución de backup y se amortice la inversión inicial, dejándola fuera del desarrollo de este proyecto y por consiguiente, de este documento.

Con la realización del estudio y el análisis de las tecnologías actuales en TI, se ha intentado destacar la gran importancia que tiene la continuidad de negocio para las empresas en el mercado actual, así como los puntos principales a tratar para conseguir la mejor solución posible en cada entorno.

El estudio preliminar fue crucial para tener una mayor visión de todas las posibilidades existentes en la actualidad. Además, la experiencia que me proporcionó la primera fase de practicas en el sector, que me permitió conocer de primera mano las diversas soluciones existentes, así como la ayuda de la propia empresa con su larga experiencia en el sector, se llegó a una de las mejores soluciones posibles. Consecuentemente, el proyecto se implantará en un futuro inmediato, llevando este trabajo a una oportunidad de negocio real.

En cuanto a la elección de la solución, a parte de las características mencionadas a lo largo del trabajo para hacer la selección final, se tuvo en cuenta los costes aproximados de los propios equipos, aunque estos no se detallen en ningún punto del trabajo. Esto es debido a la gran variación que sufren los costes a lo largo del año por parte de los propios fabricantes. Por tanto, una vez se pase a la fase de implementación real del proyecto se calcularán los costes reales, así como los detalles de facturación a los clientes.

Se ha podido comprobar lo importante y esencial que es marcarse unos objetivos claros a la hora de afrontar un proyecto de estas características, así como establecer unos requisitos mínimos, según las necesidades del cliente final, con el fin de lograr un buen producto y así satisfacer todas sus necesidades.

Por último, se ha de destacar la importancia de dedicar el tiempo que sea necesario a la fase de análisis preliminar, con el objetivo de sentar las bases y visualizar todo el desarrollo del proyecto, fase por fase, para llegar a un buen resultado final.

6 LÍNEAS FUTURAS

Como se ha ido comentando durante el desarrollo del proyecto, se dejan abiertas varias líneas futuras para continuar el desarrollo de esta plataforma, como las mencionadas a continuación:

- El primer paso una vez realizado este trabajo es el realizar una búsqueda de clientes que estén interesados en un servicio de estas características, para poder así afrontar la inversión inicial.

- Una vez encontrado un primer cliente, se llevaría a cabo la fase de implementación y despliegue de la solución, así como continuar en la búsqueda de clientes que rentabilicen la inversión.

- El siguiente punto sería la implantación de un nuevo servicio de recuperación ante desastres, de forma análoga a la realizada en este proyecto y su posterior despliegue. Esto permitiría ampliar el espectro de clientes potenciales, así como ofrecer una solución adicional a los actuales clientes.

7 BIBLIOGRAFÍA

[1] Service-Level Agreement. SLA. Definición [Online].

URL: http://en.wikipedia.org/wiki/Service-level_agreement

[2] Multi-tenant. Definición [Online].

URLs: http://www.zerto.com/blog/dr-to-the-cloud/what-is-multi-tenancy-and-why-is-it-important-for-cloud-disaster-recovery/

http://smoothspan.wordpress.com/2007/10/28/multitenancy-can-have-a-161-cost-advantage-over-single-tenant/

[3] EMC Corporation. Documentación [Online].

URLs: http://spain.emc.com/index.htm?fromGlobalSiteSelect

http://www.powerlink.emc.com

http://www.emc.com/data-protection/avamar.htm

http://www.emc.com/domains/datadomain/index.htm?locationID=6

http://spain.emc.com/collateral/hardware/data-sheet/h6811-datadomain-ds.pdf [PDF]

[4] VMware. Documentación [Online].

URLs: http://www.vmware.com

https://my.vmware.com/web/vmware/login

http://www.vmware.com/files/pdf/practical_guide_bcdr_vmb.pdf

[5] Quantum Corporation. Documentación [Online].

URLs: http://www.quantum.com/

https://alliance.quantum.com/na/

[6] Symantec Corporation. Documentación [Online].

URLs: http://www.symantec.com/index.jsp

[7] EspacioRack. Empresa de housing y hosting. [Online]

URL: http://www.espaciorack.com

[8] ESG. Enterprise Strategy Group.

URL: http://www.esg-global.com

[9] Plan de recuperación antes desastres. Sans [Online PDF]

URL:http://www.sans.org/reading-room/whitepapers/recovery/disaster-recovery-plan-testing-cycle-plan-plan-cycle-563

[10] Continuidad de Negocio. [Online]

URL: http://es.wikipedia.org/wiki/Plan_de_continuidad_del_negocio

[11] Copia de seguridad. Definición [Online]

URL: http://es.wikipedia.org/wiki/Backup

[12] Centro de proceso de datos. Definición [Online]

URL: http://es.wikipedia.org/wiki/Centro_de_proceso_de_datos

[13] Hewlett-Packard. [Online]

URL:http://www8.hp.com/us/en/products/data-storage/data-storage-products.html?compURI=1225909#.U4zCwhaZbkw

Este documento esta firmado por

	**Firmante**	CN=tfgm.fi.upm.es, OU=CCFI, O=Facultad de Informatica - UPM, C=ES
	Fecha/Hora	Tue Jun 03 22:55:32 CEST 2014
	Emisor del Certificado	EMAILADDRESS=camanager@fi.upm.es, CN=CA Facultad de Informatica, O=Facultad de Informatica - UPM, C=ES
	Numero de Serie	630
	Metodo	urn:adobe.com:Adobe.PPKLite:adbe.pkcs7.sha1 (Adobe Signature)